



INSTITUT FÜR STATISTIK
SONDERFORSCHUNGSBEREICH 386



Wagenpfeil:

Dynamische Modelle zur Ereignisanalyse

Sonderforschungsbereich 386, Paper 32 (1996)

Online unter: <http://epub.ub.uni-muenchen.de/>

Projektpartner



Dynamische Modelle zur Ereignisanalyse

Inaugural-Dissertation
zur Erlangung des Grades Doctor oeconomiae publicae (Dr.oec.publ.)
an der Ludwig-Maximilians-Universität München

vorgelegt von

Stefan Wagenpfeil

1995

Referent:
Koreferent:
Promotionsabschlußberatung:

Prof. Dr. L. Fahrmeir
Prof. Dr. H. Schneeweiß
28. Februar 1996

Vorwort

Die vorliegende Arbeit entstand an der Ludwig-Maximilians-Universität München während meiner Tätigkeit als DFG-Mitarbeiter am Institut für Statistik und als Mitarbeiter am SFB 386 „Statistische Analyse diskreter Strukturen“. Die zugrunde liegende wissenschaftliche Projektarbeit wurde zu einem wesentlichen Teil von der Deutschen Forschungsgemeinschaft finanziell gefördert.

Ganz besonders danke ich Herrn Prof. Dr. L. Fahrmeir für seine großzügige Unterstützung, Förderung und Betreuung meiner Arbeit. Von ihm ging die Initiative zu diesem Forschungsprojekt aus.

Herrn Prof. Dr. H. Schneeweiß, Herrn Prof. Dr. L. Knüsel und Herrn Prof. R. Rees, Ph.D., danke ich für ihre Bereitschaft zur Erstellung der Gutachten bzw. Übernahme von Prüferaufgaben.

Ferner möchte ich mich bei meinen Kollegen Frau Dr. Rita Künstler und Herrn Christian Gieger bedanken für Korrekturlesen und bei Herrn Holger Dittmar für seine Mitarbeit beim Aufbereiten der Arbeitslosendaten und bei den Programmdurchläufen.

Für ihr Verständnis und ihre Unterstützung danke ich nicht zuletzt meiner Lebensgefährtin Gudrun Hakel.

München, 6. Oktober 1995

Stefan Wagenpfeil

Inhaltsverzeichnis

Notation und Bezeichnungen	3
1 Einleitung	5
1.1 Datensätze	7
1.1.1 Magenkrebsdaten.....	7
1.1.2 Efron-Daten	9
1.1.3 Veteran's Administration Lungenkrebsdaten	10
1.1.4 Gliomdaten.....	11
1.1.5 SOEP Arbeitslosendaten.....	13
1.2 Fragestellung und Inhaltsübersicht.....	20
2 Ereignisanalyse	24
2.1 Ein-Episoden-Ein-Zustands-Fall	24
2.2 Mehr-Zustands-Fall.....	28
2.3 Mehr-Episoden-Fall	29
2.4 Zensierung	33
2.5 Datenkodierung und Einbezug von Kovariablen	34
3 Dynamische Modelle	39
3.1 Multinomiales Zustandsraummodell	44
3.2 Piecewise exponential Modell.....	47
4 Posteriori-Modus-Schätzung in dynamischen Modellen	50
4.1 Ein-Episoden-Fall	58
4.2 Mehr-Episoden-Fall	67
5 Schätzalgorithmen	72
5.1 Schätzalgorithmen für Modelle mit informativer Start-Priori.....	72
5.1.1 Linear gewichteter Kalman Filter und Glätter (LGKFG).....	74
5.1.2 Verallgemeinerter und erweiterter Kalman Filter und Glätter (EKFG)	76
5.1.3 Iterativ gewichteter Kalman Filter und Glätter (IGKFG)	78
5.1.4 EM-Typ Algorithmus für informative Start-Priori	79

5.2	Schätzalgorithmen für Modelle mit diffuser Start-Priori	81
5.2.1	Diffuser linear gewichteter Kalman Filter und Glätter (D-LGKFG).....	89
5.2.2	Diffuser verallgemeinerter und erweiterter Kalman Filter und Glätter (D-EKFG)	91
5.2.3	Diffuser iterativ gewichteter Kalman Filter und Glätter (D-IGKFG)	92
5.2.4	EM-Typ Algorithmus für diffuse Start-Priori	93
5.2.5	EM-Typ Algorithmus für unterschiedliche Zeitintervallängen	95
5.3	Numerische Aspekte	97
6	Datenanalyse	100
6.1	Magenkrebsdaten	100
6.1.1	Dynamisches binomiales Logit-Modell	101
6.1.2	Piecewise exponential Modell	111
6.2	Efron-Daten	122
6.3	Veteran’s Administration Lungenkrebsdaten	126
6.4	Gliomdaten	132
6.5	SOEP Arbeitslosendaten	140
6.5.1	Ein-Episoden-Ein-Zustands-Analyse	140
6.5.2	Mehr-Zustands-Analyse	148
6.5.3	Mehr-Episoden-Analyse.....	154
7	Zusammenfassung und Ausblick	161
	Anhang	163
A	Programmbeschreibung zu PME.ST	163
B	Beweise	170
	Literaturverzeichnis	174

Notation und Bezeichnungen

Algorithmen

- KFG: Abkürzung für „Kalman Filter und Glätter“.
- LGKFG: Abkürzung für „linear gewichteter Kalman Filter und Glätter“.
- EKFG: Abkürzung für „verallgemeinerter und erweiterter Kalman Filter und Glätter“.
- IGKFG: Abkürzung für „iterativ gewichteter Kalman Filter und Glätter“.
- D-LGKFG: Abkürzung für „diffuser linear gewichteter Kalman Filter und Glätter“.
- D-EKFG: Abkürzung für „diffuser verallgemeinerter und erweiterter Kalman Filter und Glätter“.
- D-IGKFG: Abkürzung für „diffuser iterativ gewichteter Kalman Filter und Glätter“.

Sonstiges

- pr steht für „probability“. Mit dem Wahrscheinlichkeitsraum (Ω, \wp, pr) und $E \in \wp$ bezeichnet $\text{pr}(E)$ die Wahrscheinlichkeit für das Eintreten des Ereignisses E .
- Sei T eine Zufallsvariable. Dann steht $E(T)$, wenn nicht anders bezeichnet, für den Erwartungswert der Zufallsvariable T . $T \sim N$ bedeutet, daß T normalverteilt, $T \sim M$, daß T multinomialverteilt ist und $T \sim P$, daß T poissonverteilt ist.
- Vektoren sind grundsätzlich Spaltenvektoren.
- $\mathbb{N} = \{1, 2, 3, \dots\}$ ist die Menge der natürlichen und \mathbb{R} die Menge der reellen Zahlen.
- Mit $n \in \mathbb{N}$ bezeichnet $n!$ die Fakultät von n , also $n! := \prod_{i=1}^n i$.
- Sei A eine Menge, dann bezeichnet $|A|$ die Mächtigkeit von A . Ist A eine endliche Menge, dann gibt $|A|$ die Anzahl der Elemente von A an.
- $[,.)$ bezeichnet ein links abgeschlossenes und rechts offenes Intervall. Seien $a, b \in \mathbb{R}$. Dann ist $[a, b) = \{x \in \mathbb{R} \mid a \leq x < b\}$.
- $(,,)$ bezeichnet ein offenes Intervall. Seien $a, b \in \mathbb{R}$. Dann ist $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$.
- Mit $A \in \mathbb{R}^{n,m}$, $n, m \in \mathbb{N}$, bedeutet $A' \in \mathbb{R}^{m,n}$ die transponierte Matrix und bei $m = 1$ gilt entsprechendes für Vektoren.
- Mit $A \in \mathbb{R}^{n,n}$, $n \in \mathbb{N}$ und A regulär bezeichnet A^{-1} die inverse Matrix.
- Wenn nicht anders bezeichnet, dann bedeutet $I_n \in \mathbb{R}^{n,n}$ die Einheitsmatrix,

$$I_n := \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & 1 \end{pmatrix}, \quad n \in \mathbb{N}.$$

- Wenn nicht anders vermerkt, dann bezeichnet $e_i \in \mathbb{R}^n$ mit $i \in \{1, \dots, n\}$ den i -ten kanonischen Einheitsvektor, das heißt $e_i = (0, \dots, 0, 1, 0, \dots, 0)' \in \mathbb{R}^n$ mit der 1 an der i -ten Position, $0_n \in \mathbb{R}^n$ den Nullvektor, das heißt $0_n = (0, \dots, 0)' \in \mathbb{R}^n$ und $0_{n,n} \in \mathbb{R}^{n,n}$ die Nullmatrix, das heißt

$$0_{n,n} := \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix}, \quad \text{mit } n \in \mathbb{N}.$$

- Sei $x = (x_1, \dots, x_n)' \in \mathbb{R}^n$ und seien $h_i: \mathbb{R}^n \rightarrow \mathbb{R}$, $i = 1, \dots, m$, und $h: \mathbb{R}^n \rightarrow \mathbb{R}^m$ mit $h(x) = (h_1(x), \dots, h_m(x))'$ auf ihrem gesamten Definitionsbereich zweimal differenzierbare Abbildungen, dann bezeichnet

$$\frac{\partial h(\bar{x})}{\partial x} := \begin{pmatrix} \frac{\partial h_1(\bar{x})}{\partial x_1} & \cdots & \frac{\partial h_m(\bar{x})}{\partial x_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial h_1(\bar{x})}{\partial x_n} & \cdots & \frac{\partial h_m(\bar{x})}{\partial x_n} \end{pmatrix} \in \mathbb{R}^{n,m}$$

die Jacobi-Matrix von h ausgewertet an der Stelle $\bar{x} \in \mathbb{R}^n$, und

$$\frac{\partial^2 h_i(\bar{x})}{\partial x \partial x'} := \begin{pmatrix} \frac{\partial^2 h_i(\bar{x})}{\partial x_1 \partial x_1} & \cdots & \frac{\partial^2 h_i(\bar{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 h_i(\bar{x})}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 h_i(\bar{x})}{\partial x_n \partial x_n} \end{pmatrix} \in \mathbb{R}^{n,n}$$

die Hesse-Matrix von h_i , $i = 1, \dots, m$, an der Stelle \bar{x} .

1 Einleitung

Ereignisorientierte Datenstrukturen sind dadurch gekennzeichnet, daß sie für jede Person oder Untersuchungseinheit über die „Zeitdauern bis zu einem Zustandswechsel beziehungsweise bis zum Eintreffen bestimmter Ereignisse und über deren Abfolge“ (Blossfeld, Hamerle, Mayer, 1986, S. 5) berichten. Dabei kann man zwischen dem zeitstetigen Fall, bei dem der exakte Zeitpunkt eines Zustandswechsels dokumentiert ist, und dem zeitdiskreten Fall unterscheiden, bei dem für das Eintreten eines bestimmten Ereignisses lediglich ein Zeitintervall angegeben und zwischen aufeinanderfolgenden Zustandswechseln nur die Anzahl der Zeitintervalle bekannt ist. Diese Arbeit konzentriert sich in bezug auf Daten- und Beispielauswahl und dementsprechend bei der Modellierung und Schätzung auf den zeitdiskreten Fall, weil dieser in vielen Anwendungen, Datensätzen und Datenerhebungsschemata konkret vorliegt. Für eine ausführliche Behandlung der zeitstetigen Situation sei auf Blossfeld et al. (1986) und Cox und Oakes (1984) verwiesen, ferner auf Lancaster (1990) mit ökonomischem Schwerpunkt, auf Kalbfleisch und Prentice (1980) mit speziellem Augenmerk auf die Biometrie und mit Anwendungsbeispielen aus der Medizin und für den reinen Survival-Fall auf Fleming und Harrington (1991) mit der Zählprozeß-Sichtweise.

Ereignisdaten treten in den verschiedensten Bereichen des öffentlichen und privaten Lebens auf und werden wiederum in unterschiedlichen wissenschaftlichen Gebieten und empirisch forschenden Disziplinen wie den Wirtschafts- und Sozialwissenschaften, der Materialkunde, den Ingenieurwissenschaften, der Medizin und der Psychologie analysiert. Aus diesem Grund ist die Terminologie und Notation sehr unterschiedlich. Ist die Dauer der von der jeweiligen Untersuchungseinheit in einem Zustand verbrachten Zeit gegeben, so spricht man, je nach Anwendungsgebiet, von Verweildauer, Lebensdauer, Verschleißdauer, Episodendauer, Wartezeit oder Ausfallzeit. Unabhängig von den unterschiedlichen Begriffen in den verschiedenen Wissenschaftszweigen kann man sich vom mathematischen Standpunkt aus die Datenlage und Datenentstehung in der Ereignisanalyse als stochastischen Prozeß mit endlichem Zustandsraum in stetiger oder, wie in unserem Fall, diskreter Zeit vorstellen. Im Hinblick auf die Modellierung und Schätzung lassen sich Ereignisdaten in drei Gruppen einteilen, auf die in dieser Arbeit stets Bezug genommen wird:

Man spricht von Ein-Episoden-Ein-Zustands-Daten, wenn nur eine Episode betrachtet und lediglich ein Zielzustand berücksichtigt wird. Unter einer Episode versteht man dabei den

Zeitabschnitt zwischen zwei aufeinanderfolgenden Zustandswechseln. Den in dieser Zeit von der Untersuchungseinheit eingenommenen Zustand bezeichnet man als transient. Der Ziel- beziehungsweise Endzustand ist der Zustand, in den nach Ablauf der Episode gewechselt wird. Diese Datenlage wird im biometrischen Kontext auch als Survival-Fall bezeichnet und es gibt für sie zwei mögliche Entstehungsursachen:

- 1.) Der Zustandswechsel erfolgt in einen absorbierenden Zielzustand. Damit ist gemeint, daß der Zielzustand, einmal eingenommen, nicht mehr verlassen werden kann. Ein Beispiel wäre der Tod eines Patienten in einer Lebensdauerstudie.
- 2.) Die Untersuchungseinheit wechselt zwar im Laufe der Zeit in verschiedene Zustände, wobei sich die Zustände auch wiederholen können, zum Beispiel die verschiedenen Arbeitslosigkeitsepisoden eines Arbeitnehmers innerhalb eines fest vorgegebenen Beobachtungszeitraumes. Gleichzeitig ist aber nur die Verweildauer in einem dieser Zustände von Interesse, unabhängig von dem Zustand, zu dem anschließend gewechselt wird. Dies wäre beispielsweise der Fall, wenn man lediglich an der ersten Arbeitslosigkeitsepisode des Arbeitnehmers interessiert wäre.

Betrachtet man ebenfalls nur eine Episode und existieren mehrere (absorbierende) Zielzustände, so liegt der Mehr-Zustands-Fall vor. Ein Beispiel dafür wäre die Beobachtung der Verweildauer im Zustand Leben und, wenn der Tod eintritt, die Unterscheidung nach verschiedenen Todesursachen. Manche Autoren wie zum Beispiel Hamerle und Tutz (1989) und Courgeau und Lelièvre (1992) bezeichnen diese Situation auch als Competing-Risks-Fall, während beispielsweise Cox und Oakes (1984) diesen Begriff in einem anderen Sinn verwenden. Eine Diskussion über die verschiedenen in der Literatur vorgestellten Competing-Risks-Betrachtungsweisen und Modellansätze findet sich in Kalbfleisch und Prentice (1980, Abschnitt 7.2).

Sind außerdem im Verlauf der Zeit mehrere Übergänge von einem Zustand zum anderen möglich oder tritt ein Ereignis wiederholt auf, dann handelt es sich um Mehr-Episoden-Daten.

Ein wichtiges und häufig anzutreffendes Phänomen im Bereich der Ereignisanalyse ist ferner die Zensierung. Von Rechtszensierung spricht man, wenn das Verweilen einer Untersuchungseinheit in einem bestimmten Zustand nicht bis zu Ende, das heißt bis zum unmittelbar darauffolgenden Zustandswechsel, beobachtet oder festgestellt werden konnte. Im Rahmen

einer Lebenszeitstudie beispielsweise wären hierfür verschiedene Ursachen denkbar, etwa der Umzug einer Untersuchungsperson, so daß sie für eine Weiterverfolgung nicht mehr zur Verfügung steht und keine weiteren Daten von ihr erhoben werden können. Ein weiterer Grund könnte darin liegen, daß der interessierende Zustandswechsel nicht innerhalb des fest vorgegebenen Beobachtungszeitraumes der Studie stattfindet.

Für die Modellierung der Ereignisdaten und auch bei der Schätzung wird in dieser Arbeit das Konzept der Hazardrate benutzt, das in Kapitel 2 kurz skizziert und beispielsweise in Fahrmeir und Tutz (1994, Kapitel 9) und in Kalbfleisch und Prentice (1980) näher erläutert ist.

Beispiele für die verschiedenen, in der Ereignisanalyse auftretenden Datensituationen geben die im nachfolgenden Abschnitt vorgestellten Datensätze. Die Zielsetzung dieser Arbeit, die Ergebnisse, die gewonnen wurden, und eine Inhaltsübersicht folgen in Abschnitt 1.2.

1.1 Datensätze

Anhand des folgenden Datenmaterials sollen die in dieser Arbeit dargestellten Modelle zur statistischen Analyse motiviert, die vorgeschlagenen Schätzalgorithmen und -verfahren auf ihre Leistung hin überprüft und ihre Eigenschaften illustriert werden. Überdies handelt es sich meistens um bekannte Datensätze, die in der Literatur häufig zitiert werden und damit Referenzcharakter besitzen. Damit können zusätzlich die hier erhaltenen Ergebnisse mit den in der Literatur angegebenen Resultaten verglichen, eingeschätzt und bewertet werden. Auf die jeweiligen Datensätze wird an verschiedenen Stellen des Textes in Form von Beispielen immer wieder Bezug genommen.

1.1.1 Magenkrebsdaten

Dieser Datensatz aus dem medizinischen Bereich, der in Gamerman (1991) abgedruckt ist, enthält die Überlebenszeiten von 90 an Magenkrebs erkrankten Patienten. Die Verweildauer im Zustand „Leben“ ist in Tagen und damit in diskreter Form angegeben. Der absorbierende End- beziehungsweise Zielzustand ist der „Tod“. Weitere Abgangsarten sind nicht dokumentiert. Somit haben die Daten die Ein-Episoden-Ein-Zustands-Struktur. Nicht bei allen Patienten konnte jedoch der Tag beobachtet werden, an dem sie vom Zustand „Leben“ in den Zustand „Tod“ wechselten. Diese sind als zensiert gekennzeichnet. Außerdem wurden

die Patienten zufällig in zwei gleich große Behandlungsgruppen aufgeteilt. Die erste Gruppe wurde sowohl mit Chemo- als auch mit Strahlentherapie behandelt, während die zweite lediglich der Chemotherapie unterzogen wurde. In der ersten Behandlungsgruppe beträgt die größte beobachtete Überlebenszeit 1736 Tage, wobei 7 Zensierungen vorliegen, und in der zweiten 1694 Tage, wobei 3 Verweildauerzeiten zensiert sind. Über die Verweildauer von vier ausgewählten Patienten aus diesem Datensatz gibt Abbildung 1.1 Auskunft.

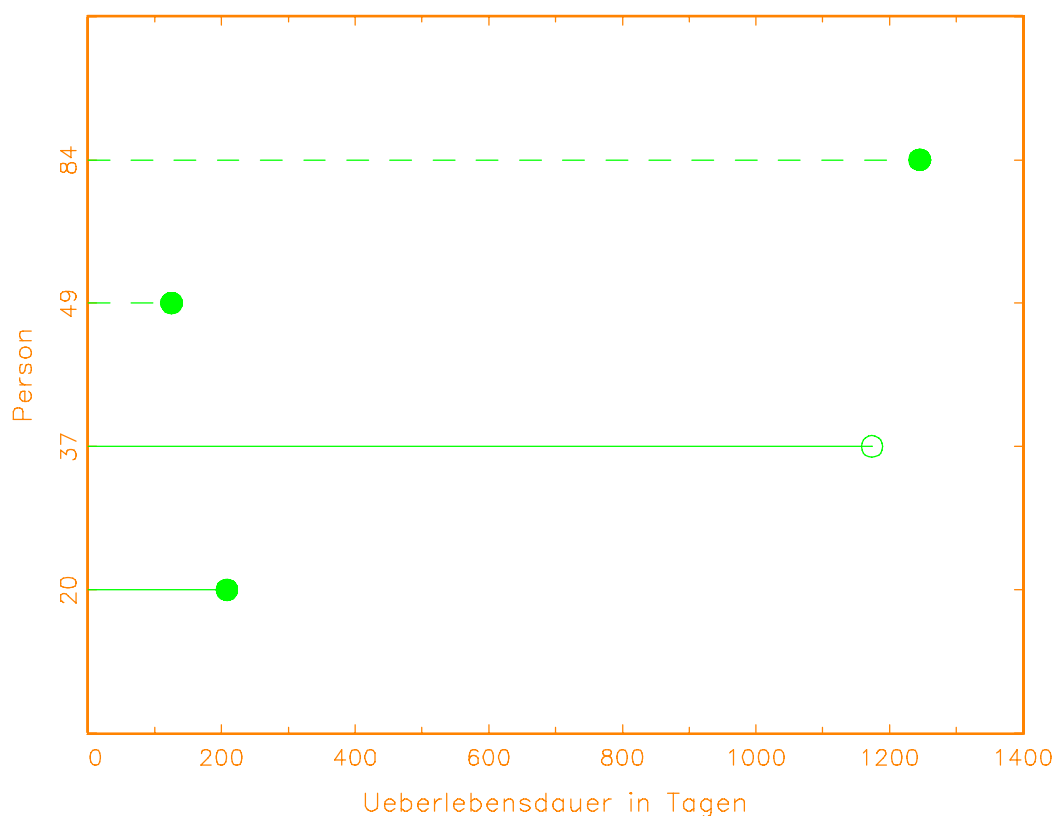


Abbildung 1.1 : Verweildauer von vier ausgewählten Patienten aus dem Magenkrebs-Datensatz: — Chemo- und Strahlentherapie, - - - Chemotherapie, • Ereignis „Tod“ ist eingetreten, o Zensierung. Die Personennummer bezieht sich dabei auf die Reihenfolge im Datensatz.

Gegenstand der Analyse und Untersuchung dieses Datensatzes ist die Beantwortung der Frage, welche der beiden Behandlungsmethoden, die kombinierte Chemo- und Strahlenthe-

rapie oder die Anwendung der Chemotherapie alleine, für die Patienten die bessere Therapie wäre. Als besser soll dabei die Methode gelten, die, im Vergleich zur anderen, ein geringeres „Sterberisiko“ und längere Überlebenszeiten mit sich bringt. Ferner wäre zu klären, ob sich eine der beiden Therapien ständig als die bessere erweist, oder ob im Laufe der Beobachtungszeit Veränderungen auftreten, die, abhängig von der Verweildauerzeit, einmal die eine, dann die andere Behandlungsart als die bessere ausweisen.

1.1.2 Efron-Daten

In diesem Datensatz, der aus einer klinischen Studie zur Behandlung von Drüsenkrebs stammt und Efron (1988) entnommen ist, sind die Lebensdauern von 96 an Drüsenkrebs erkrankten Personen in Tagen festgehalten. Der einzig vorhandene und gleichzeitig absorbierende Endzustand ist wie bei den Magenkrebsdaten der „Tod“. Somit haben die Daten auch hier die Ein-Episoden-Ein-Zustands-Struktur: eine Episode, nämlich die Verweildauer im Zustand „Leben“ und ein Abgangszustand, der Übergang in den Zustand „Tod“. Die Patienten, zu unterschiedlichen Zeiten in die Studie aufgenommen, wurden wiederum zufällig in zwei alternative Behandlungsgruppen eingeteilt. Die erste Gruppe mit 51 Personen erhielt nur eine Strahlentherapie. Bei der zweiten Behandlungsgruppe bestehend aus den übrigen 45 Patienten wurde die Strahlentherapie mit einer Chemotherapie gekoppelt. In der ersten Therapiegruppe beträgt die größte beobachtete Überlebenszeit 1417 Tage bei 9 Zensierungen und in der zweiten 2297 Tage bei 14 Zensierungen. Weitere Kovariable sind in diesem Datensatz nicht enthalten. Die Verweildauern von vier ausgewählten Patienten sind in Abbildung 1.2 dargestellt.

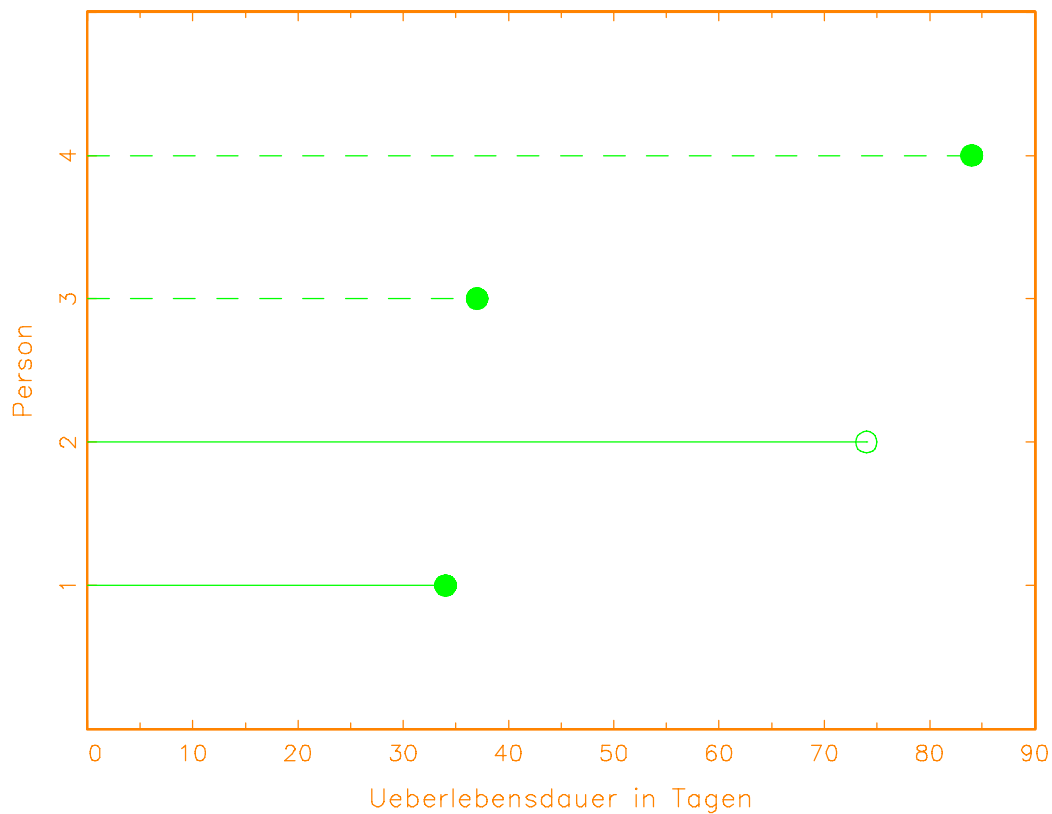


Abbildung 1.2 : Verweildauer von vier ausgewählten Patienten aus dem Efron-Datensatz:
 — Chemo- und Strahlentherapie, — — — Chemotherapie, ● Ereignis „Tod“ ist eingetreten, ○ Zensierung.

Wie bei den Magenkrebsdaten steht auch hier die Frage im Vordergrund, welche der beiden medizinischen Therapien, im Vergleich zur jeweils anderen, die bessere im Bezug auf Sterberisiko und Überlebenszeit wäre.

1.1.3 Veteran's Administration Lungenkrebsdaten

Diese Daten wurden von der Veteran's Administration in einer klinischen Studie zur Behandlung von Lungenkrebs erhoben. Es nahmen insgesamt 137 männliche Patienten mit fortgeschrittenem, nicht mehr operierbarem Lungenkrebs teil und der Datensatz enthält ihre Überlebenszeit in Tagen. Der einzig vorhandene, absorbierende Zielzustand war der „Tod“ und somit liegt auch hier die Ein-Episoden-Ein-Zustands-Situation vor. Die Studiendauer,

die zu Beginn nicht festgelegt war, richtete sich nach den Überlebensdauern der Personen. Dies ist ein möglicher Grund für die vergleichsweise geringe Anzahl von neun zensierten Überlebenszeiten. Die Patienten wurden ebenfalls, wie in den vorhergehenden beiden Beispielen, zufällig in zwei Therapiegruppen eingeteilt. Die erste Gruppe erhielt die klassische, die zweite Gruppe eine neue, zu testende Chemotherapie. Um die Heterogenität innerhalb des Patientenkollektivs zu berücksichtigen, wurden weitere Kovariablen erfaßt:

x_K : Allgemeinzustand nach dem Karnofsky-Index in drei Kategorien: 10-30 vollkommen bettlägrig, 40-60 teilweise bettlägrig, 70-90 kann selbst für sich sorgen.

x_M : Anzahl der Monate zwischen Diagnose und Eintritt in die Studie.

x_A : Alter in Jahren.

x_{Th} : Frühere Therapie in zwei Kategorien: ja, nein.

x_{Tu} : Histologischer Tumortyp in vier Kategorien: schuppig, klein, drüsenartig, groß.

x_G : Therapiegruppe in zwei Kategorien: klassische Chemotherapie, zu testende Chemotherapie.

Der Datensatz zusammen mit einer genauen Beschreibung der Kovariablen ist in Kalbfleisch und Prentice (1980, Kap. 3.8.2 und Appendix 1) enthalten.

Neben der Frage, welche der beiden Arten von Chemotherapien die bezüglich Überlebenszeit bessere ist, können hier noch die Einflüsse der Kovariablen auf das Sterberisiko einzeln oder in Kombination untersucht und berücksichtigt werden.

1.1.4 Gliomdaten

Die Deutsche Hirntumorgruppe hat eine Multicenter-Studie mit einer Gesamtlaufzeit von sieben Jahren durchgeführt und dabei 501 an malignem Hirntumor leidende Patienten rekrutiert. Die Überlebenszeit wurde in Tagen notiert und das absorbierende Zielereignis war „Tod“. Die Datenstruktur ist wieder vom Ein-Episoden-Ein-Zustands-Typ.

Nach einer Tumoroperation wurden die Patienten zufällig in zwei Behandlungsgruppen aufgeteilt. Die erste Gruppe erhielt neben einer Strahlenbehandlung auch eine Chemothe-

rapie mit der Bezeichnung BCNU, während die zweite Gruppe zusätzlich zur Strahlentherapie eine kombinierte Chemotherapie BCNU + VM26 bekam. Ferner wurden viele, den Krankheitszustand und -verlauf der Patienten dokumentierende prognostische Faktoren als Kovariablen erhoben. Sie sind in Ulm et al. (1989) aufgeführt.

Im Rahmen dieser Arbeit wurde ein Teildatensatz aus der Studie verwendet, der auf der Begleitdiskette zur Diplomarbeit von Aydemir (1992) als ASCII-Datei gespeichert ist. Er wird im folgenden als Gliomdatensatz bezeichnet und enthält die vollständigen Daten von 467 Patienten. Neben ihren Überlebenszeiten in Tagen sind auch die folgenden Kovariablen festgehalten:

x_K : Allgemeinzustand nach dem Karnofsky-Index in drei Kategorien: ≤ 60 , 70-80, ≥ 90 .

Je größer der Karnofsky-Index, desto besser ist der Gesundheitszustand.

x_M : Malignität in drei Kategorien: Grad 3, Grad 4, Grad 3-4.

Die Malignität gibt die Bösartigkeit des Tumors an, wobei ein zunehmender Malignitätsgrad einer größeren Bösartigkeit entspricht.

x_A : Alter in zwei Kategorien: ≤ 50 Jahre, > 50 Jahre.

x_G : Therapiegruppe in zwei Kategorien: Strahlentherapie und BCNU Chemotherapie, Strahlentherapie und kombinierte Chemotherapie BCNU + VM26.

x_R : Resektion in zwei Kategorien: biopsiert, reseziert.

Resektion ist der chirurgische Fachbegriff für die Entfernung (in diesem Falle des Tumors). Biopsie bezeichnet dabei die Entnahme von Tumorgewebstücken. Dies erfolgt in der Regel bei Patienten, deren Krankheitszustand eine Operation nicht mehr zulässt. Reseziert bedeutet die teilweise operative Entfernung des Tumors.

x_E : Epileptische Anfälle in zwei Kategorien: ja, nein.

Die längste beobachtete Verweildauer war 2220 Tage, insgesamt sind 92 Überlebenszeiten zensiert.

Anhand dieses Datensatzes stellen sich die folgenden Fragen, deren Beantwortung Ziel einer statistischen Analyse sein kann:

- Wie groß ist der Einfluß der Kovariablen auf die Überlebenszeit beziehungsweise das Sterberisiko?
- Liegt eine dynamische Veränderung ihres Einflusses vor, das heißt ändert sich der Effekt der Kovariablen mit der Zeit?
- Welchen möglicherweise zeitabhängigen Einfluß hat die Therapie bei einer Subgruppenanalyse, wenn also zum Beispiel nur eine bestimmte Altersklasse berücksichtigt wird?
- Welche Resultate ergeben sich bei einem Vergleich zwischen dem Therapieeffekt in einer Subgruppenanalyse und dem Therapieeffekt im Rahmen einer Gesamtanalyse, bei der das komplette Patientenkollektiv einbezogen ist?

1.1.5 SOEP Arbeitslosendaten

Das Sozio-ökonomische Panel (SOEP) ist eine umfassende Längsschnittuntersuchung von privaten Haushalten in Deutschland. Die Befragung wird jährlich seit 1984 durchgeführt. Dabei werden, soweit möglich, stets dieselben Haushalte und Personen befragt. Zu den ursprünglich 5921 westdeutschen Haushalten mit 12290 Personen kamen im Juni 1990 weitere 2179 Haushalte mit 4453 Personen aus den neuen Bundesländern hinzu. Das Projekt ist am Deutschen Institut für Wirtschaftsforschung (DIW) in Berlin angesiedelt und aus einem Teilprojekt des von der Deutschen Forschungsgemeinschaft (DFG) geförderten Sonderforschungsbereiches 3 „Mikroanalytische Grundlagen der Gesellschaftspolitik“ hervorgegangen. Weitere Auskünfte über das SOEP und seine Entwicklung von 1984 bis 1994 finden sich in dem Aufsatz der Projektgruppe Sozio-ökonomisches Panel (1995).

Die Daten des SOEP enthalten eine Vielzahl von Informationen über die Lebensbedingungen und die Lebensqualität der befragten Menschen wie Einkommensverläufe, Gesundheit, Lebenszufriedenheit, Weiterbildung und Qualifikation. Im Rahmen dieser Arbeit konzentrieren wir uns jedoch auf die Daten des SOEP, die in Zusammenhang mit der Arbeitslosigkeit stehen und bezeichnen diese Teildatensätze als SOEP Arbeitslosendaten. Im folgenden wird kurz skizziert, wie diese Daten im Sinne der Ereignisanalyse sowohl für Ein-Episoden-Ein-Zustands-, für Mehr-Zustands- und auch für Mehr-Episoden-Untersuchungen benutzt werden können.

In jedem Jahr von 1984 bis 1993 wurden die interviewten Personen unter anderem nach ihrem Erwerbsstatus befragt. Einen entsprechenden Auszug aus den Fragebögen für das Jahre 1984 zeigt die Abbildung 1.3:

Abbildung 1.3: Auszug aus dem SOEP Fragebogen für die Welle B. Quelle: Deutsches Institut für Wirtschaftsforschung (Hrsg.) (1993).

Die Befragung, die 1985 stattfand, war dabei retrospektiv auf die Situation im unmittelbar vorhergehenden Jahr, also auf 1984, bezogen. Damit ist für jede Person, die an den Panel-Befragungen von 1984 bis 1993 teilgenommen hat, ihr Erwerbsstatus monatsweise von Januar 1983 bis Dezember 1992 nach den neun, in Abbildung 1.3 dargestellten Kategorien bekannt. In Abbildung 1.4 ist der Verlauf des Erwerbsstatus von drei ausgewählten Befragten aufgezeichnet, wobei die Kategorien der Erwerbstätigkeit folgendermaßen kodiert sind:

- 0: keine Angabe beziehungsweise missing value
- 1: vollerwerbstätig

- 2: Teilzeitarbeit oder nebenberufliche Tätigkeit
- 3: betriebliche Ausbildung, Fortbildung oder Umschulung
- 4: arbeitslos
- 5: Rente oder Pension
- 6: Schule oder Hochschule
- 7: Wehr- oder Zivildienst
- 8: Hausmann oder Hausfrau
- 9: Sonstiges

Die Nummern der Personen beziehen sich auf die (zufällige) Reihenfolge im verwendeten Teildatensatz und stehen in keinem Zusammenhang mit den Personennummern, die im Rahmen des SOEP vergeben werden.

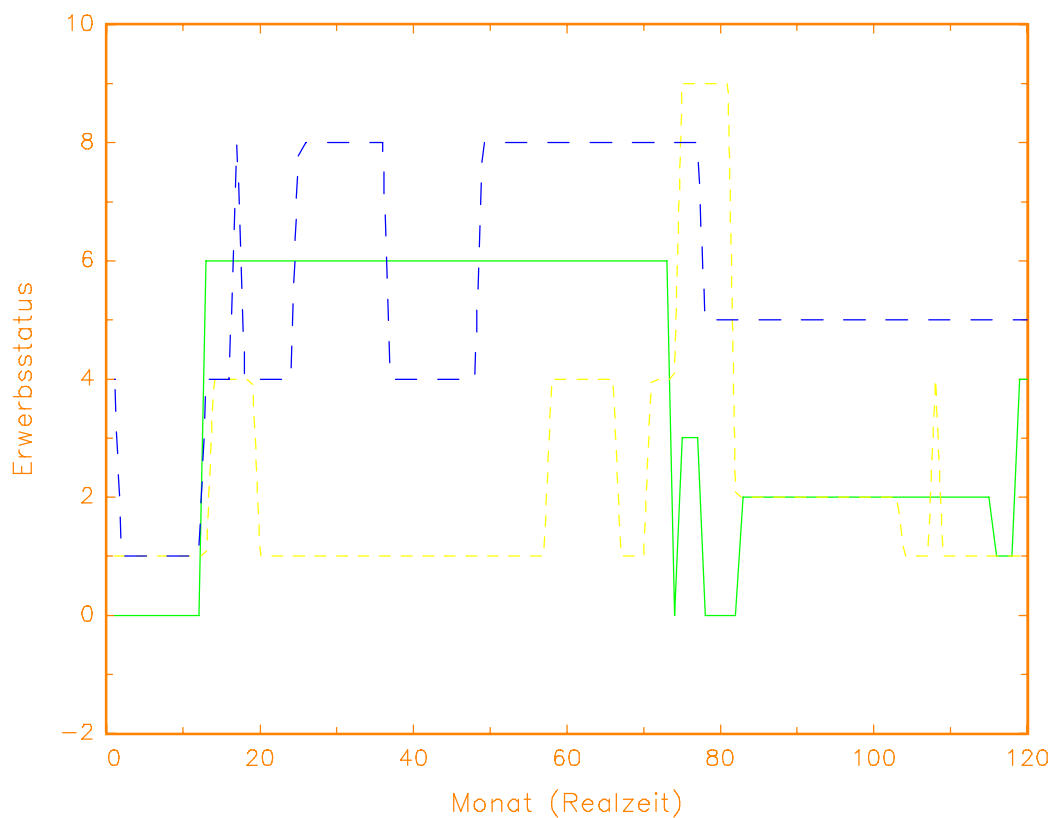


Abbildung 1.4: Erwerbsstatus der — Person 1, - - - Person 9 und - - - - Person 19 von Januar 1983 bis Dezember 1992, das sind 120 Monate.

Von der Person mit der Nummer 1 beispielsweise ist der Erwerbsstatus im ersten Jahr nicht bekannt. Im 13. Monat vom Januar 1983 an gerechnet, das ist der Januar 1984, wechselt sie in den Zustand 6, der für Schule oder Hochschule steht. Sechzig Monate lang, bis zum Dezember 1988, bleibt sie in diesem Zustand, bevor sie über mehr oder weniger kurze Episoden in den Zuständen Umschulung, Teilzeittätigkeit und Vollzeitarbeit im Monat 119, das ist der November 1992, arbeitslos wird. Ob die Arbeitslosigkeit auch nach dem 120. Monat noch anhält, ist nicht ersichtlich, weil der Beobachtungszeitraum im Dezember 1992 endet. Daher liegt hier eine rechtszensierte Arbeitslosigkeitsepisode vor. Auf dieselbe Weise lassen sich die Erwerbsverläufe der beiden anderen Personen ablesen. Insbesondere fällt bei Person 9 auf, daß sie zu Beginn des Beobachtungszeitraumes, also im Januar 1983, bereits arbeitslos war. Eine solche Situation bezeichnet man als linkszensiert.

Zur Analyse der Arbeitslosigkeitsdauer und ihrer Einflußfaktoren genügt es, nur die Daten der Personen zu betrachten, die im Beobachtungszeitraum von Januar 1983 bis Dezember 1992 mindestens einmal arbeitslos waren. In Abbildung 1.5 sind die Arbeitslosigkeitsepisoden der Personen aus Abbildung 1.4 in der Realzeit, das ist die Kalenderzeit in Monaten gerechnet ab Januar 1983, aufgetragen.

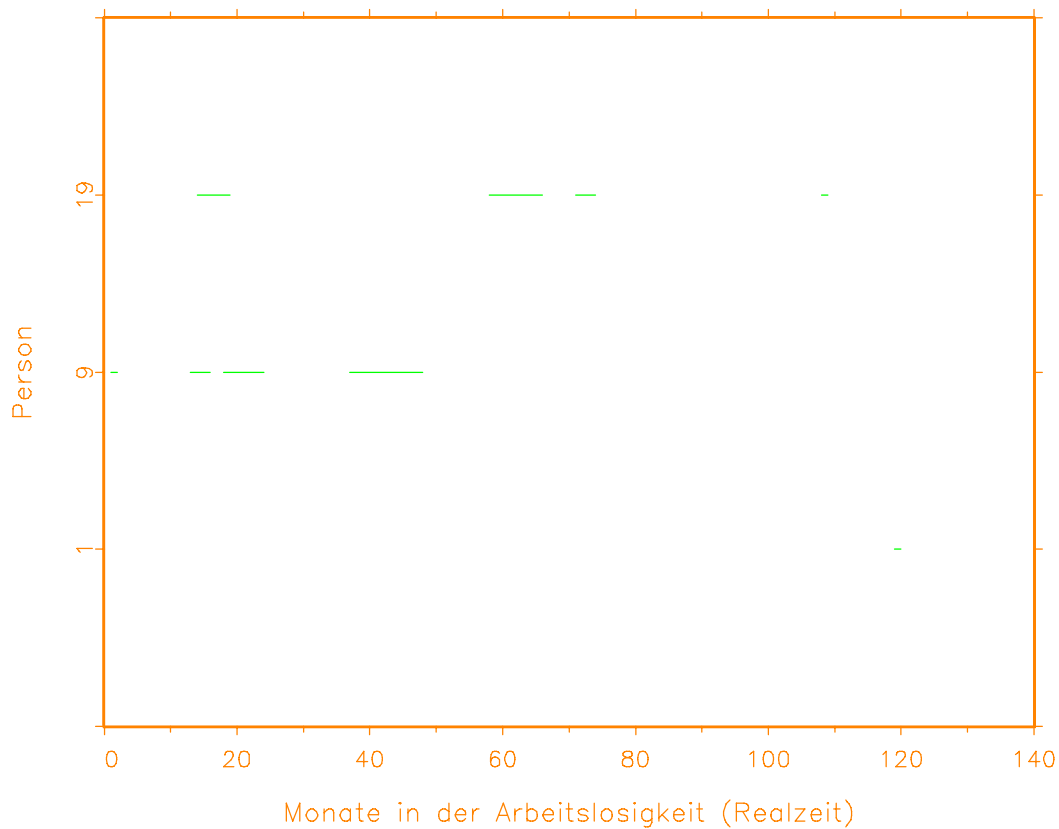


Abbildung 1.5 Arbeitslosigkeitsepisoden der Personen mit den SOEP-Nummern 1, 9 und 19.

Person 1 hat nur eine (zensierte) Arbeitslosigkeitsepisode erlebt, während die Personen 9 und 19 viermal arbeitslos waren.

Im Rahmen der Verweildaueranalyse interessiert man sich nicht mehr für die Real- beziehungsweise Kalenderzeit, sondern für die Verweildauerzeit in einem bestimmten Zustand. Die Zeitrechnung beginnt dann für jede Untersuchungseinheit mit dem Eintritt in den betrachteten Zustand. Sie endet, wenn der betreffende Zustand, in unserem Falle die Arbeitslosigkeit, wieder verlassen wird oder wenn Rechtszensierung vorliegt. Für eine Ein-Episoden-Ein-Zustands-Analyse wären beispielsweise die Verweildauern in der ersten Arbeitslosigkeitsepisode geeignet. Diese sind für die drei Personen mit den Nummern 1, 9 und 19 in der Abbildung 1.6 dargestellt.

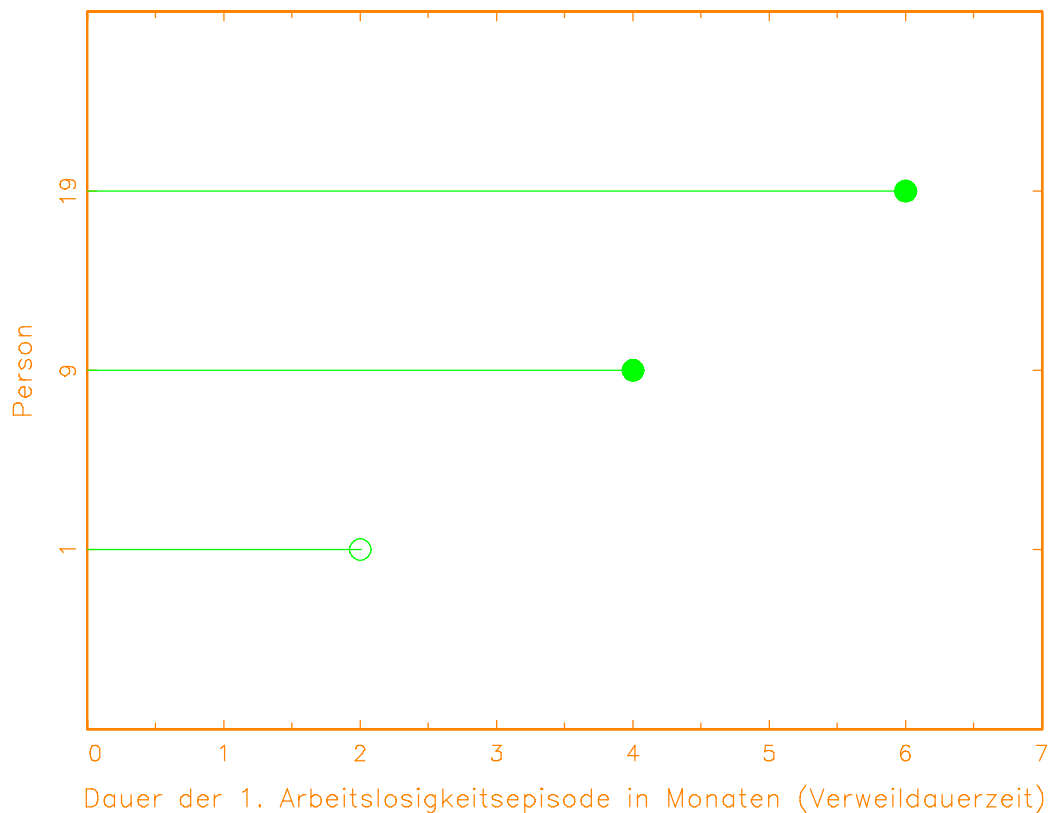


Abbildung 1.6 Verweildauern in der ersten Arbeitslosigkeitsepisode: • Ereignis eingetreten, das heißt Austritt aus der Arbeitslosigkeit, o Zensurierung.

Vergleicht man die Abbildungen 1.5 und 1.6, so bezieht sich die angegebene Verweildauerzeit bei Person Nummer 1 auf die (zensierte) Arbeitslosigkeit in den Monaten 119 und 120. Da nicht bekannt ist, wie lange vor dem Januar 1983 die Person mit der Nummer 9 schon arbeitslos war, wird ihre linkszensierte Arbeitslosigkeitsepisode im ersten Monat nicht berücksichtigt und stattdessen die Arbeitslosigkeit vom 13. bis 16. Monat, das ist von Januar 1984 bis April 1984, als erste Arbeitslosigkeitsepisode betrachtet. Bei der Person mit der Nummer 19 liegt die erste Arbeitslosigkeit in den Monaten 14 bis 19, das ist von Februar 1984 bis Juli 1984.

Im Rahmen einer Mehr-Zustands-Analyse ist man nicht nur an der Dauer der einen betrachteten Episode interessiert, sondern auch daran, in welchen von mehreren möglichen absorbierenden Zielzuständen anschließend gewechselt wird. In unserem Beispiel soll also nicht

nur berücksichtigt werden, wann der Austritt aus der ersten Arbeitslosigkeitsepisode stattfindet, sondern auch welche Form der Erwerbstätigkeit sich unmittelbar anschließt. Aus diesem Grund sind in der Abbildung 1.7 für unsere Personen 1, 9 und 19 zusätzlich zu den Verweildauern in der Arbeitslosigkeit auch die Abgangszustände aufgetragen. Die Kodierung der Erwerbstätigkeit ist dabei analog zu Abbildung 1.4 .

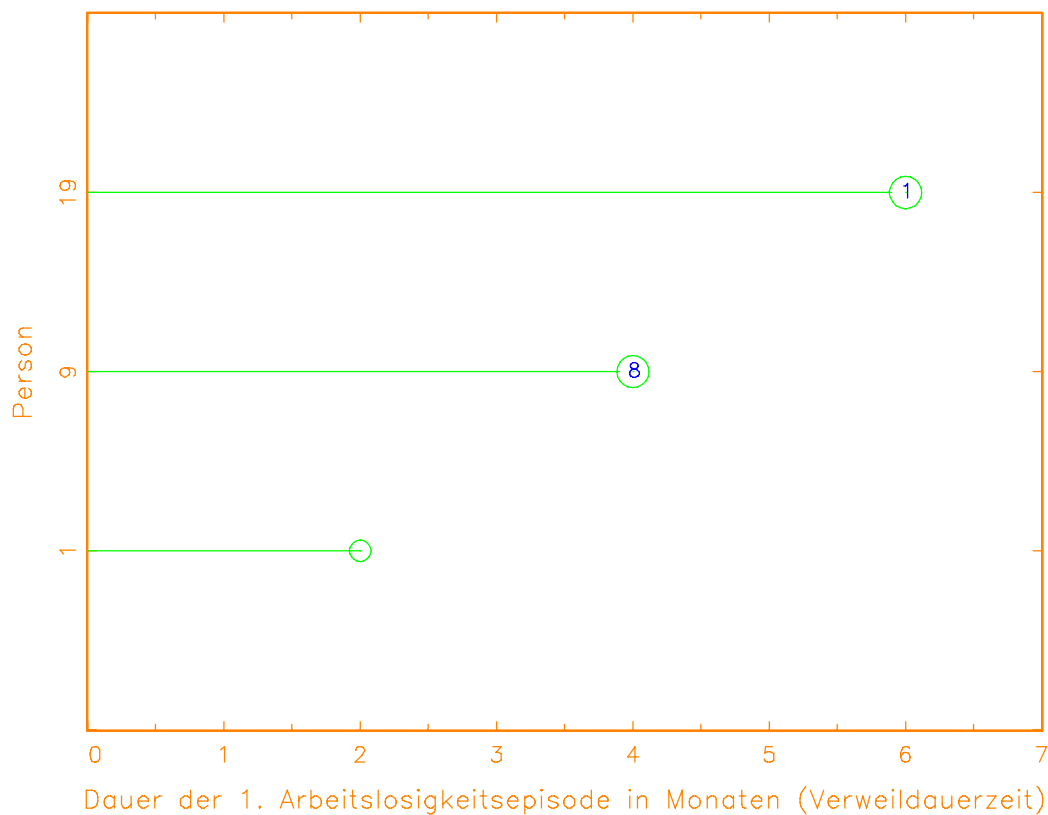


Abbildung 1.7 Verweildauern in der ersten Arbeitslosigkeitsepisode und Abgangszustände:
1: Wechsel in eine Vollzeittätigkeit, 8: Hausmann oder Hausfrau, o Zensurierung.

Der dazugehörige Datensatz, der sich sowohl für eine Ein-Episoden-Ein-Zustands- als auch für eine Mehr-Zustands-Analyse eignet, enthält von jeder Person, die im Zeitraum von Januar 1983 bis Dezember 1992 an der SOEP-Befragung teilgenommen hat und in dieser

Zeit mindestens einmal arbeitslos war - insgesamt waren das 2911 Personen - , die Verweildauer in der ersten Arbeitslosigkeitsepisode und den Abgangszustand. Außerdem wurden noch die drei Kovariablen Nationalität (in drei Kategorien), Geschlecht (in zwei Kategorien) und Alter (vierkategorial) aufgenommen.

Damit können Hazardrate und Kovariableneffekte abgangsspezifisch untersucht werden. Außerdem lassen sich Vergleiche zwischen den einzelnen Abgangsarten vornehmen. Dadurch können verglichen mit der Ein-Zustands-Analyse zusätzliche Aufschlüsse und Erkenntnisse über den anhand des Datensatzes zu untersuchenden Gegenstand, in unserem Fall die Arbeitslosigkeit, gewonnen werden.

Für eine Mehr-Episoden-Untersuchung wurde ein weiterer Datensatz aufbereitet, der von allen Personen, die im Zeitraum von Januar 1983 bis Dezember 1992 an der Befragung teilgenommen haben, die zweite Arbeitslosigkeitsepisode enthält, und zwar sowohl die Verweildauer, gerechnet ab Eintritt in die zweite Arbeitslosigkeit, als auch den Abgangszustand, kodiert wie in Abbildung 1.4. Zusätzlich zu den Kovariablen Nationalität, Geschlecht und Alter, die wie im ersten Teildatensatz kategorisiert sind, wurden noch die Dauer der ersten Arbeitslosigkeit und die Zeit zwischen der ersten und der zweiten Arbeitslosigkeitsepisode, jeweils gemessen in Monaten, aufgenommen. Insgesamt sind die Daten von 816 Personen berücksichtigt.

Mit ihm läßt sich auch die zweite Arbeitslosigkeit in Abhängigkeit von der Dauer der ersten zusammen mit den anderen, im Datensatz enthaltenen Kovariablen analysieren.

1.2 Fragestellung und Inhaltsübersicht

In der vorliegenden Arbeit werden Methoden und Verfahren zur Analyse zeitdiskreter Ereignisdaten im Rahmen von dynamischen Modellen vorgestellt. Besondere Gesichtspunkte sind dabei die simultane Schätzung von Hazardrate und zeitabhängigen Kovariableneffekten. Die exogenen Variablen beziehungsweise prognostischen Faktoren können ebenfalls über die Zeit hinweg variieren. Ferner wird besprochen, wie die Rechtszensierung in die Modellierung und Schätzung einbezogen werden kann. Weitere Ergebnisse dieser Arbeit können wie folgt zusammengefaßt werden:

- Zustandsraummodelle bilden eine flexible und vielfältige Grundlage für die Untersuchung und Analyse von diskreten Ereignisdaten.

- Der Kalman Filter und Glätter ist ein effizientes Verfahren zur Lösung spezieller blocktri-diagonaler linearer Gleichungssysteme hoher Dimension. Damit können auf numerisch günstigem Wege Fisher-Scoring-Schritte durchgeführt werden. Außerdem stehen am Ende der Berechnung die Hauptdiagonalblöcke der inversen Fisher-Informationsmatrix zur Verfügung und müssen nicht erst durch zusätzlichen rechnerischen Aufwand bestimmt werden.
- Der diffuse Kalman Filter und Glätter erlaubt es, Modelle mit nichtinformativer Start-Priori-Verteilung zu verwenden. Das hier entwickelte Verfahren benötigt dabei keinen höheren numerischen Aufwand als der klassische lineare Kalman Filter und Glätter und läßt sich auf kanonischem Wege von diesem ableiten. Außerdem können häufig auftretende Probleme bei der Startwertwahl der Hyperparameter vermieden werden.

In Kapitel 2 wird die Struktur von Ereignisdaten formal dargestellt und es werden für die folgenden Kapitel wichtige Variable und Begriffe definiert. Nach einem Überblick über verschiedene Modellansätze und Betrachtungsweisen in der Ereignisanalyse folgt ein Abschnitt über Zensierung. Im letzten Abschnitt dieses Kapitels wird der Einbezug von Kovariablen besprochen und eine zweite Art, die Daten zu kodieren. Dabei gehen wir auch auf das Konzept der Hazardrate ein.

Zustandsraummodelle, die man auch als dynamische Modelle bezeichnet, werden in Kapitel 3 besprochen. Sie enthalten zwei Gruppen von zu schätzenden Parametern: einerseits dynamische Regressionskoeffizienten, die in diesem Zusammenhang auch als Zustände bezeichnet werden, und andererseits Hyperparameter. Wie man Zustandsraummodelle im Rahmen der Ereignisanalyse einsetzen kann, wird in Abschnitt 3.1 gezeigt. Modelle für zeitdiskrete Beobachtungen lassen sich auch durch Diskretisierung stetiger Modelle erstellen. Ein Beispiel hierfür ist das Piecewise exponential Modell, das in Abschnitt 3.2 in Zustandsraumform gebracht wird. Ein weiterer Vorteil von Zustandsraummodellen ist ihre Allgemeinheit. Liegt für diese Modellklasse ein effizientes Schätzverfahren vor, dann können Schätzungen und Analysen im Rahmen all der Modelle durchgeführt werden, die Zustandsraumform haben.

In Kapitel 4 wird ein knapper Überblick gegeben über die verschiedenen Schätzkonzepte, die für Zustandsraummodelle verwendet werden können. Anschließend wird eines dieser Konzepte, die Posteriori-Modus-Schätzung, die auf der Optimierung eines penalisierten Loglikelihood-Kriteriums basiert, ausführlich besprochen. Ferner wird gezeigt, wie und

unter welchen Voraussetzungen diese Schätzmethodik im Ein-Episoden-Fall (Abschnitt 4.1) und im Mehr-Episoden-Fall (Abschnitt 4.2) angewendet werden kann.

Die Schätzalgorithmen selbst, ihre Herleitung, Motivation und Eigenschaften sind Gegenstand von Kapitel 5. Zunächst wird, über verschiedene Zwischenschritte, ein Verfahren zur Lösung des penalisierten Likelihood-Kriteriums mit informativer Start-Priori-Verteilung bei gegebenen Hyperparametern besprochen. Zur Schätzung der Hyperparameter wird der in Fahrmeir (1994) vorgestellte EM-Typ Algorithmus angegeben. Weil jedoch, mit Ausnahme von simulierten Daten, bei realen Datensätzen im allgemeinen keine Information über die Start-Priori-Verteilung vorliegt, wird in Abschnitt 5.2 das Verfahren für eine diffuse Start-Priori explizit hergeleitet. Anschließend modifizieren wir den zuvor dargestellten EM-Typ Algorithmus zur Hyperparameterschätzung für die Situationen einer diffusen Start-Priori-Verteilung und für den Fall von unterschiedlichen Zeitintervalllängen. Numerische Aspekte, die insbesondere bei der Implementierung von Bedeutung sind, werden in Abschnitt 5.3 angesprochen. All diese Schätzverfahren sind in dem Programmpaket PME.ST implementiert, das in GAUSS programmiert ist und vom Autor angefordert werden kann.

Mit den zuvor beschriebenen Verfahren und Methoden werden in Kapitel 6 die Daten, die in der Einführung vorgestellt wurden, analysiert. Bei den Magenkrebsdaten wird unter anderem eine Subgruppenanalyse, das ist die nach den einzelnen Therapiegruppen getrennt durchgeführte Schätzung, mit einer Gesamtanalyse verglichen. Auch der Gebrauch von unterschiedlichen Zeitintervalllängen und „Zeitgrids“ im Rahmen des Piecewise exponential Modells wird untersucht. Anhand der Efron-Daten vergleichen wir die Resultate, die wir unter der Anwendung von zwei verschiedenen Modellen erhalten. Sowohl bei den Veteran's Administration Lungenkrebsdaten aus Teilabschnitt 1.1.3 als auch bei den Gliomdaten aus Teilabschnitt 1.1.4 werden mehrere Kovariable gleichzeitig in die Untersuchung einbezogen. Anhand der SOEP Arbeitslosendaten wird gezeigt, daß eine Mehr-Zustands-Analyse zusätzliche Erkenntnisse im Vergleich zur Ein-Zustands-Analyse bringen kann. Im Rahmen einer Mehr-Episoden-Analyse wird auch die zweite Arbeitslosigkeitsepisode untersucht. Dadurch erhält man Aufschlüsse über den weiteren Verlauf nach der ersten Arbeitslosigkeit. Kapitel 7 faßt die bisher dargestellten methodischen und empirischen Resultate zusammen und gibt einen Ausblick auf die Punkte, die Gegenstand weiterer Forschungsaktivitäten sein können.

Der Anhang A enthält eine tabellarische Programmbeschreibung zu PME.ST, und in Anhang B finden sich die technischen Beweise, auf die im Text verwiesen wird.

Fast alle Grafiken in dieser Arbeit wurden mit dem Statistik-Programmpaket GAUSS erstellt. Eine Ausnahme bildet die Abbildung 2.1 in Kapitel 2, die unter Verwendung von Microsoft EXCEL 5.0 angefertigt wurde.

2 Ereignisanalyse

Im folgenden werden für die späteren Kapitel wichtige Variablen und Begriffe vorgestellt und die Datenlage in der Ereignisanalyse formal dargestellt. Dies erfolgt zunächst getrennt nach den drei Gruppen von Ereignisdaten. Nach einem kurzen Abschnitt über die Zensierung werden die Notation und Indexvariablen eingeführt, die bei der Motivation und Beschreibung der Modelle und Schätzprinzipien verwendet werden.

2.1 Ein-Episoden-Ein Zustands-Fall

Zunächst betrachten wir den Ein-Episoden-Ein-Zustands-Fall, das heißt man zeichnet die Verweildauer T einer Untersuchungseinheit U in einem transierten Zustand J_U auf. Bei Eintritt des zu beobachtenden Ereignisses wechselt U in einen absorbierenden Zustand.

Einen Überblick über Entwicklung und Grundgedanken der verschiedenen in der Literatur besprochenen Modellkonzepte gibt Singpurwalla (1995). Er teilt sie nach ihren jeweiligen Modellansatzpunkten in vier Gruppen ein:

- Stochastische Modelle, die den Zustand J_U von U beschreiben. Er verweist dabei auf die Arbeiten von Wenocur (1989) und Lemoine und Wenocur (1985).
- Beschreibung sowohl des Zustands J_U von U als auch den einer Kovariablen, die U beeinflusst, durch einen Markovschen additiven Prozeß. Diese stochastischen Prozesse wurden von Çinlar (1972) eingeführt.
- Modellierung der Ausfall- beziehungsweise Hazardrate (das heißt letztendlich der Verteilung von T) durch einen stochastischen Prozeß. Eine Referenz hierfür ist die Arbeit von Arjas (1981), die den Modellansatz aus der Zählprozeßsichtweise der „skandinavischen Schule“ beschreibt.
- Modellierung der vermuteten Einflußgrößen der Hazardrate durch einen stochastischen Prozeß. Dabei unterscheidet Singpurwalla zwischen dem Fall,
 - 1.) bei dem den zeitvariierenden Kovariablen und dem Fall,
 - 2.) bei dem den dynamischen Kovariableneffekten ein stochastischer Prozeß zugrundegelegt wird.

Im ersten Fall wird durch den angenommenen Kovariablenprozeß ein Hazardratenprozeß induziert. Das Vorgehen unter Punkt 2.) wollen wir als das Konzept der Hazardrate

bezeichnen. Singpurwalla (1995) verweist dazu auf das Piecewise exponential Modell in Gamerman (1991), dessen Darstellung das Schätzprinzip der konjugierten Priori von West, Harrison und Migon (1985) im Rahmen von dynamischen generalisierten linearen Modellen verwendet.

Der Modellansatz, der in dieser Arbeit verfolgt und im nächsten Kapitel dargestellt wird, läßt sich ebenfalls unter Punkt 2.) einordnen. Allerdings wird zur Schätzung eine andere Idee als in Gamerman (1991) benutzt: der Posteriori-Modus-Ansatz, der in Kapitel 4 besprochen wird.

Sei im folgenden T die Zufallsvariable Verweildauer mit $T \geq 0$. Die Hazardrate zur Zeit $T=t$ (gemeint ist die Realisierung der Zufallsvariablen T mit Wert t) ist definiert als

$$\lambda(t) := \frac{f(t)}{S(t)}, \quad (2.1)$$

wobei $S(t) := \text{pr}(T \geq t) = 1 - F(t)$ die Survivor-Funktion und $F(t)$ die Verteilungsfunktion von T ist. Für $f(t)$ unterscheiden wir zwei Fälle:

- Falls T eine (absolut) stetige Zufallsvariable ist, dann ist $f(t)$ die Dichte von T .
- Falls T eine diskrete Zufallsvariable ist, dann ist $f(t)$ die Wahrscheinlichkeitsfunktion von T .

Im folgenden werden die Struktur und die Eigenschaften der Hazardrate getrennt nach beiden Fällen dargestellt.

Zunächst sei T eine (absolut) stetige Zufallsvariable. $f(t)$ bezeichnet die Dichte von T und es gilt daher $f(t) = dF(t)/dt = \lim_{\substack{\Delta t \rightarrow 0 \\ \Delta t > 0}} \{ \text{pr}(t \leq T < t + \Delta t) / \Delta t \}$.

Unter Verwendung von (2.1) ergibt sich

$$\lambda(t) = \lim_{\substack{\Delta t \rightarrow 0 \\ \Delta t > 0}} \frac{\text{pr}(t \leq T < t + \Delta t | T \geq t)}{\Delta t}, \quad (2.2)$$

das heißt die Hazardrate ist der Grenzwert einer bedingten Wahrscheinlichkeit. Sie läßt sich als augenblickliche Abgangsrate aus dem transienten in den absorbierenden Zustand zur Zeit t interpretieren, vorausgesetzt die Untersuchungseinheit U befindet sich mindestens bis zur Zeit t im transienten Zustand. Aus (2.1) ergibt sich:

- $\lambda(t) \geq 0$.

- $\lambda(t)$ ist keine Wahrscheinlichkeit, da Werte größer als Eins angenommen werden können.
- $\lambda(t) = -\{d \log S(t)/dt\}$. Verwendung von $S(0)=1$ und Integration ergibt

$$S(t) = \exp\left(-\int_0^t \lambda(s) ds\right). \quad (2.3)$$

Damit ist durch die Angabe der Hazardrate bereits die komplette Verteilung von T festgelegt. Insbesondere gilt: Hat man eine Schätzung für die Hazardrate, so liegt über (2.3) auch gleichzeitig ein Schätzer für $S(t)$ vor.

Weil $T \geq 0$ ist, läßt sich wiederum mit $S(t)$ der Erwartungswert von T in kompakter Weise schreiben, wobei wir annehmen: $\exists b > 0, b \in \mathbb{R}$ mit $F(b) = 1$. Es gilt dann, siehe dazu auch Kendall, Stuart und Ord (1987, Seite 69, Aufgabe 2.24):

$$E(T) = \int_0^b S(t) dt. \quad (2.4)$$

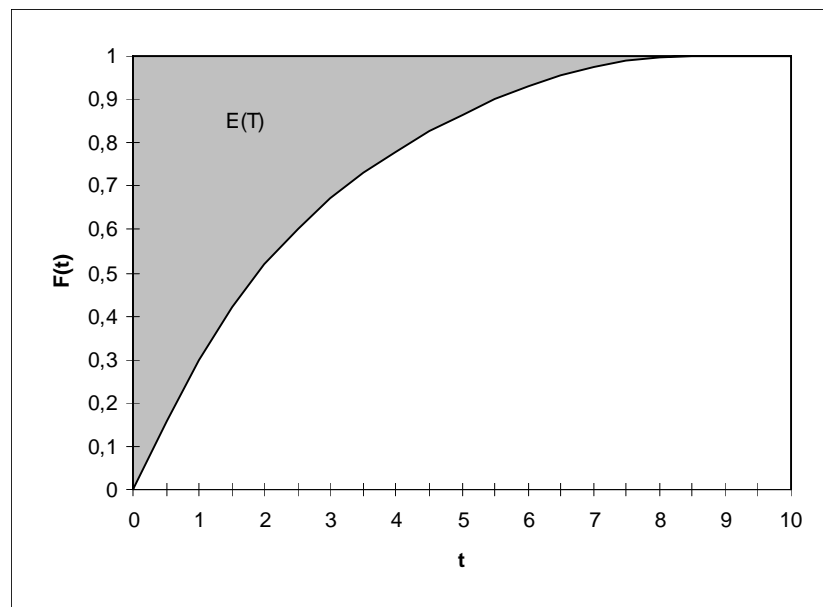


Abbildung 2.1: Darstellung des Erwartungswertes einer nichtnegativen Zufallsvariable als Fläche zwischen dem Graphen der Verteilungsfunktion $F(t)$ und der Linie mit Ordinatenwert 1 für $b = 8,5$. Die Fläche ist hier grau gepunktet.

Die Abbildung 2.1 gibt eine anschauliche Deutung für die Formel in (2.4) mit $b = 8,5$. Ein formaler Beweis findet sich in Anhang B. Mit (2.4) läßt sich der Erwartungswert von T einfach berechnen, falls das Integral auf der rechten Seite leicht auszuwerten ist.

Oftmals läßt sich der Zeitpunkt eines Zustandswechsels nicht exakt angeben oder beobachten, sondern nur ein Zeitintervall, in dem das Ereignis stattgefunden haben muß. Solche Daten bezeichnet man als intervallzensiert. Die Tatsache, daß der Zustandswechsel von mehreren Untersuchungseinheiten möglicherweise in ein und demselben Zeitintervall erfolgt und damit sogenannte „ties“ auftreten, kann bei zeitstetigen Modellen zu Problemen führen. Daher wird im folgenden der zeitdiskrete Fall betrachtet.

Sei jetzt T eine diskrete Zufallsvariable mit Wertemenge $\{1, \dots, \bar{T}\} \subseteq \mathbb{N}$. Ferner sei die Zeitachse in $\bar{T} + 1$ Intervalle $[b_0, b_1), [b_1, b_2), \dots, [b_{\bar{T}-1}, b_{\bar{T}}), [b_{\bar{T}}, \infty)$ mit $b_0 \leq b_1 \leq b_2 \leq \dots \leq b_{\bar{T}}$ eingeteilt. Oft wird $b_0 = 0$ angenommen. $T=t$ bedeutet dann, daß der Zustandswechsel im Intervall $[b_{t-1}, b_t)$ stattgefunden hat. $f(t)$ ist hier die Wahrscheinlichkeitsfunktion, das heißt $f(t) = \text{pr}(T = t)$. Mit (2.1) ergibt sich

$$\lambda(t) = \text{pr}(T = t | T \geq t), \quad t = 1, \dots, \bar{T}. \quad (2.5)$$

Damit ist die Hazardrate im zeitdiskreten Fall eine Wahrscheinlichkeit, nämlich die bedingte Wahrscheinlichkeit dafür, daß der Zustandswechsel im Intervall mit der Nummer t stattfindet, vorausgesetzt die Untersuchungseinheit U verweilt mindestens bis t im transienten Zustand Z_U . Die Wahrscheinlichkeit für das Eintreten eines Ereignisses beziehungsweise eines Zustandswechsels im Intervall Nummer t ergibt sich sofort aus (2.1) zu

$$\text{pr}(T = t) = f(t) = \lambda(t) \cdot S(t). \quad (2.6)$$

Die Survivor-Funktion kann auch hier in Abhängigkeit von der Hazardrate dargestellt werden:

$$S(t) = \prod_{i=1}^{t-1} (1 - \lambda(i)). \quad (2.7)$$

Die Beweise für (2.7) finden sich in Anhang B. Nach (2.7) ist alleine durch die Angabe der Hazardrate die Verteilung der diskreten Zufallsvariablen T bestimmt.

2.2 Mehr-Zustands-Fall

Im folgenden Abschnitt wird wieder nur eine Episode betrachtet, aber der Zustandswechsel soll nach $j \in \{1, \dots, \bar{J}\}$, einen von $\bar{J} \in \mathbb{N}$ möglichen absorbierenden Endzuständen, erfolgen. J bezeichne die diskrete Zufallsvariable Endzustand. In der Literatur lassen sich zwei Arten von Ansatzpunkten für eine Modellbildung finden.

Die eine Möglichkeit besteht darin, für jede Untersuchungseinheit und für jeden der \bar{J} möglichen Endzustände latente Verweildauern $T_1, \dots, T_{\bar{J}}$ anzunehmen. T_j ist dann die theoretische Verweildauer im transienten Zustand bis der Zustandswechsel nach j stattfindet. Die anderen Abgangsarten $1, 2, \dots, j-1, j+1, \dots, \bar{J}$ werden nicht berücksichtigt. Beobachten kann man jedoch nur die Zufallsvariablen (T, J) mit $T := \min\{T_1, \dots, T_{\bar{J}}\}$, $J = j$, falls $T = T_j$. Diese Darstellung bezeichnen zum Beispiel Cox und Oakes (1984) als „competing risks“, und Arjas (1989) beschreibt diesen Ansatz vom Blickwinkel der markierten Punktprozesse.

Die gemeinsame Verteilung (T, J) von T und J kann auf verschiedenste Weise spezifiziert werden. Wir gehen dabei von der nachfolgenden Darstellung aus. Sei im folgenden die Zeitachse in $\bar{T}+1$ Zeitintervalle wie in Abschnitt 2.1 unterteilt, T diskret mit Werten $t \in \{1, \dots, \bar{T}\} \subseteq \mathbb{N}$ und J wie oben. Die abgangsspezifische Hazardrate nach j , definiert durch

$$\lambda_j(t) := \text{pr}(T = t, J = j | T \geq t), \quad (2.8)$$

ist die bedingte Wahrscheinlichkeit für einen Zustandswechsel nach j im Zeitintervall Nummer t unter der Voraussetzung, daß t erreicht wurde und vorher kein Abgang in einen Endzustand erfolgt ist. Es wird dabei angenommen, daß die Untersuchungseinheit nicht gleichzeitig in zwei oder mehrere verschiedene absorbierende Zielzustände wechseln kann. Die globale Hazardrate beziehungsweise Gesamthazardrate ist definiert als

$$\lambda(t) := \sum_{j=1}^{\bar{J}} \lambda_j(t) = \text{pr}(T = t | T \geq t). \quad (2.9)$$

Sie ist die bedingte Wahrscheinlichkeit für das Eintreten eines Ereignisses (ganz gleich welches) im Intervall t , vorausgesetzt das Zeitintervall t wurde erreicht.

Die Survivor-Funktion ergibt sich in Abhängigkeit von der globalen Hazardrate wie im Ein-Zustands-Fall zu

$$S(t) = \prod_{i=1}^{t-1} (1 - \lambda(i)) . \quad (2.10)$$

Die Wahrscheinlichkeitsfunktion von T für den Übergang nach $J = j$ läßt sich ebenfalls in Abhängigkeit von der Gesamthazardrate schreiben:

$$f_j(t) := \text{pr}(T = t, J = j) = \lambda_j(t) \cdot S(t) . \quad (2.11)$$

Der Beweis zu (2.11) findet sich in Anhang B. Die Randwahrscheinlichkeitsfunktion ist dann gegeben durch

$$f(t) := \text{pr}(T=t) = \sum_{j=1}^{\bar{J}} \text{pr}(T = t, J = j) \stackrel{(2.11)}{=} \sum_{j=1}^{\bar{J}} \lambda_j(t) \cdot S(t) \stackrel{(2.9)}{=} \lambda(t) \cdot S(t) .$$

Damit ist durch $\lambda_j(t)$, $j = 1, \dots, \bar{J}$, $t = 1, \dots, \bar{T}$, die Randverteilung von T spezifiziert. Außerdem ist mit der Angabe der abgangsspezifischen Hazardrate $\lambda_j(t)$ auch die Randverteilung von J gegeben. Dies ergibt sich aus (2.9), (2.10) und

$$\text{pr}(J = j) = \sum_{t=1}^{\bar{T}} \lambda_j(t) \cdot S(t) . \quad (2.12)$$

Der Beweis von (2.12) findet sich in Anhang B.

Für $\bar{J} = 1$ reduziert sich der Mehr-Zustands-Fall auf den Ein-Episoden-Ein-Zustands-Fall in Abschnitt 2.1.

2.3 Mehr-Episoden-Fall

Nun soll die Situation vorliegen, daß jede Untersuchungseinheit mehrere Episoden in ein und demselben oder auch in verschiedenen Zuständen erleben kann. Ein Beispiel sind die verschiedenen Arbeitslosigkeitsepisoden im SOEP-Datensatz aus Teilabschnitt 1.1.5. Die Zahl der maximal beobachteten Episoden sei auf $\bar{K} \in \mathbb{N}$ begrenzt. Für $\bar{K} = 1$ ergibt sich als Spezialfall die Mehr-Zustands-Situation aus Abschnitt 2.2. Lindsey (1993, §§ 7 und 8) gibt einen Überblick über Modelle und Verfahren für Mehr-Episoden-Daten, die in der Literatur besprochen werden.

Ein möglicher Modellansatz ist die Beschreibung derartiger dynamischer Prozesse mit Hilfe von Markov-Prozessen, Semi-Markov-Prozessen oder Erneuerungsprozessen im zeitstetigen Fall, vergleiche etwa Snyder (1975, §§ 5, 6 und 7) und mit Markov-Ketten erster

oder höherer Ordnung in der zeitdiskreten Situation. Eine grundlegende und anwendungsorientierte Einführung in den Bereich der stochastischen Prozesse bietet das Buch von Fahrmeir, Kaufmann und Ost (1981). Probleme bereitet bei diesem Ansatz jedoch der Einbezug von erklärenden Variablen.

Hamerle (1985a) stellt Zählprozess-Modelle vor, die zeitabhängige Kovariable mit festen Effekten zulassen. Insbesondere werden hier Regressionsmodelle als spezielle Zählprozess-Modelle dargestellt.

Gamerman (1992) verallgemeinert das piecewise exponential Modell, das in Gamerman (1991) für Survival-Daten verwendet wird, zur Analyse von Punktprozessen. Es kann sich dabei um multivariate Punktprozesse handeln und die Unterscheidung nach mehreren Abgangsarten, die Gamerman (1992) als „competing risks“ bezeichnet, ist ebenfalls möglich. Insbesondere erlaubt das piecewise exponential Modell problemlos den Einbezug von Kovariablen, die ebenso wie deren Effekte über die Zeit variieren können.

Das gleiche gilt für das Konzept der Hazardrate, auf das wir uns konzentrieren und auch für diese Datenlage beibehalten wollen. Dazu muß die Darstellung auf den Mehr-Episoden-Fall erweitert werden. Die zeitstetige Situation wird in Hamerle (1989) behandelt. Hamerle (1985b) geht vom zeitdiskreten Fall und abgangsspezifischen, aber zeitunabhängigen Kovariableneffekten aus. In Hamerle (1986) ist die Zeitachse auch in Intervalle diskretisiert, die Modelle können aber zeitvariierende und in additiver Form individuen-spezifische (das heißt zufällige) Effekte enthalten.

Sei im folgenden die Zeitachse wieder wie in Abschnitt 2.1 in $\bar{T}+1$ Intervalle unterteilt. $b_0 = 0$ bezeichne den Beginn und \bar{T} das Ende des Beobachtungszeitraumes. Ferner sei wie in Abschnitt 2.2 \bar{J} die Anzahl der möglichen Zustände, zu denen ein Wechsel stattfinden kann. Die zufälligen Übergänge einer Untersuchungseinheit U von einer Episode in die andere können mit zwei Arten von Zufallsvariablen beschrieben werden. Erlebt U im Laufe der Beobachtungszeit $K_U \leq \bar{K}$ Zustandswechsel, so beschreiben die nicht negativen Zufallsvariablen $T_0 := b_0 = 0 \leq T_1 \leq T_2 \leq \dots \leq T_{K_U}$ mit $T_{K_U} \leq \bar{T}$ das jeweilige Zeitintervall, in dem für $k = 1, \dots, K_U$ der k -te Zustandswechsel von U stattfindet. Die zugehörigen Zufallsvariablen $J_0, J_1, J_2, \dots, J_{K_U}$ bezeichnen die Nummer des Zustandes, in den nach der k -ten Episode, $k = 1, \dots, K_U$, gewechselt wird. Zur Vereinfachung der Notation werden die Zufallsvariablen T_0, T_k, J_0, J_k , $k = 1, \dots, K_U$, nicht mit U indiziert.

Die Untersuchungseinheit befindet sich zur Zeit $T_0 = 0$ im Zustand $J_0 = j_0 \in \{1, \dots, \bar{J}\}$. Für $t \in \{1, \dots, \bar{T}\}$, $j \in \{1, \dots, \bar{J}\}$ und $k \in \{1, \dots, K_U\}$ bedeutet $(T_k = t, J_k = j)$, daß die Untersuchungseinheit, nachdem bereits $k-1$ Zustandswechsel erfolgt sind, beim k -ten Übergang, der im t -ten Zeitintervall stattfindet, vom Zustand $j_{k-1} \in \{1, \dots, \bar{J}\}$ in den Zustand j wechselt. Dabei sei vorausgesetzt, daß $j \neq j_{k-1}$. Die Vorgeschichte zur k -ten Episode sei in $H_{k-1} = \{j_0, T_1 = t_1, J_1 = j_1, \dots, T_{k-1} = t_{k-1}, J_{k-1} = j_{k-1}\}$ zusammengefaßt mit $H_0 := \{j_0\}$. Dann ist die übergangsspezifische Hazardrate der k -ten Episode gegeben durch

$$\lambda_j^k(t) := \text{pr}(T_k = t, J_k = j | T_k \geq t, H_{k-1}) . \quad (2.13)$$

Das ist die Wahrscheinlichkeit, daß der k -te Zustandswechsel für die jeweilige Untersuchungseinheit im Zeitintervall Nummer t stattfindet und in den Zustand mit der Nummer j erfolgt unter der Voraussetzung, daß der k -te Übergang frühestens zur Zeit t eintritt und die Vorgeschichte durch H_{k-1} gegeben ist. Falls t_{k-1} die Nummer des Zeitintervalls bezeichnet, in dem der $(k-1)$ -te Übergang stattfand, dann gilt:

$$\lambda_j^k(t) = 0 \quad \forall t \in \{1, \dots, t_{k-1} - 1\} . \quad (2.14)$$

Kann die Untersuchungseinheit nach Beendigung der k -ten Episode nicht in den Zustand j wechseln, weil beispielsweise aus exogenen Gründen ein Zustandswechsel von j_{k-1} nach j nicht möglich ist, so kann dieser Sachverhalt in (2.13) dadurch berücksichtigt werden, daß man

$$\lambda_j^k(t) = 0 \quad \forall t \geq t_{k-1}$$

setzt.

Für Zeitintervalle, in denen Zustandswechsel erfolgen, machen wir die gleiche Annahme wie Hamerle (1985b). Es sollen für solche Zeitintervalle die Hazardraten so gelten, als wenn sich die Untersuchungseinheit während des gesamten Zeitintervalls in dem neuen Zustand befunden hätte. Die Gesamthazardrate für die k -te Episode ist gegeben durch

$$\lambda^k(t) = \text{pr}(T_k = t | T_k \geq t, H_{k-1}) = \sum_{j=1}^{\bar{J}} \lambda_j^k(t) . \quad (2.15)$$

Die Survivor-Funktion für die k -te Episode ist die bedingte Wahrscheinlichkeit dafür, daß der k -te Zustandswechsel frühestens im Zeitintervall t auftritt, gegeben die Vorgeschichte H_{k-1} . Entsprechend läßt sich die bedingte Verteilungsfunktion der k -ten Episode

$F^k(t|H_{k-1})$, gegeben H_{k-1} , schreiben als $F^k(t|H_{k-1}) = 1 - S^k(t|H_{k-1})$. Mit den gleichen Überlegungen wie im Beweis zu (2.7) kann man zeigen, daß

$$S^k(t|H_{k-1}) := \text{pr}(T_k \geq t|H_{k-1}) = \prod_{i=1}^{t-1} (1 - \lambda^k(i)). \quad (2.16)$$

Aus (2.14) und (2.16) folgt dann

$$S^k(t|H_{k-1}) = 1 \quad \forall t \in \{1, \dots, t_{k-1} - 1\}. \quad (2.17)$$

Mit den Überlegungen wie im Beweis zu (2.11) ergibt sich, daß die bedingte Wahrscheinlichkeitsfunktion der k -ten Episode für einen Übergang nach $J_k = j$, gegeben H_{k-1} , dargestellt werden kann durch

$$f_j^k(t|H_{k-1}) := \text{pr}(T_k = t, J_k = j|H_{k-1}) = \lambda_j^k(t) \cdot S^k(t|H_{k-1}). \quad (2.18)$$

Durch Summation erhält man die bedingte Wahrscheinlichkeitsfunktion der k -ten Episode, gegeben H_{k-1} , zu

$$f^k(t|H_{k-1}) := \text{pr}(T_k = t|H_{k-1}) = \sum_{j=1}^{\bar{J}} \text{pr}(T_k = t, J_k = j|H_{k-1}) \stackrel{(2.18)}{=} \lambda^k(t) \cdot S^k(t|H_{k-1}). \quad (2.19)$$

Aus (2.15), (2.16) und (2.19) ergibt sich, daß mit der Angabe der abgangsspezifischen Hazardrate für die k -te Episode aus (2.13), $\lambda_j^k(t)$, bereits die gesamte bedingte Verteilung von T_k , gegeben die Vorgeschichte H_{k-1} , spezifiziert ist.

Ferner ist mit $\lambda_j^k(t), j = 1, \dots, \bar{J}, t = t_{k-1}, \dots, \bar{T}, T_{k-1} = t_{k-1}$, auch die bedingte Verteilung von J_k gegeben H_{k-1} festgelegt. Dies ergibt sich aus (2.15), (2.17) und

$$\text{pr}(J_k = j|H_{k-1}) = \sum_{t=t_{k-1}}^{\bar{T}} \lambda_j^k(t) \cdot S^k(t|H_{k-1}). \quad (2.20)$$

(2.20) läßt sich wie (2.12) beweisen unter Beachtung von (2.14) und unter Verwendung von (2.18).

Die bisherige Betrachtung war kalenderzeitorientiert, das heißt die Zeit t bis zum Eintritt des k -ten Zustandswechsels wurde gemessen ab b_0 , dem Beginn der Beobachtung. Dem entsprechend wurde die Hazardrate in Abhängigkeit von der Kalenderzeit t angesetzt. Ist man jedoch nicht an der (absoluten) Kalenderzeit, sondern an der Verweildauer $s = t - t_{k-1}$ mit $T_{k-1} = t_{k-1}$ interessiert, die die Untersuchungseinheit in der k -ten Episode

verbringt mit $k \leq K_U$, so läßt sich alternativ die Hazardrate $\tilde{\lambda}$ in Abhängigkeit von s spezifizieren:

$$\tilde{\lambda}_j^k(s) := \lambda_j^k(t). \quad (2.21)$$

In unseren Untersuchungen und Anwendungen nehmen wir an, daß für jede Untersuchungseinheit der Beginn der ersten Episode mit dem Beginn des Beobachtungszeitraumes b_0 zusammenfällt und somit keine linkszensierten Episoden vorliegen, oder daß die Vorgeschichte keinen Einfluß auf die Geschehnisse ab b_0 hat. Die letzte Episode K_U , die eine Untersuchungseinheit erfährt, kann jedoch zensiert sein, etwa weil der Beobachtungszeitraum fest vorgegeben ist. Dieses Problem der Rechtszensierung wird im nächsten Abschnitt angesprochen.

2.4 Zensierung

Eine Möglichkeit, rechtszensierte Daten zu behandeln, besteht darin, im Rahmen eines Mehr-Zustands-Modelles für die Zensierung eine eigene Abgangsart zu reservieren. Wir wollen jedoch in dieser Arbeit die Rechtszensierung explizit bei dem in Kapitel 4 besprochenem Posteriori-Modus-Schätzkonzept berücksichtigen. Dies geschieht dadurch, daß man die Informationen über eine zensierte Episode in die Schätzung einbezieht, die man hat: Das ist die Dauer zwischen dem letzten Zustandswechsel beziehungsweise dem Beginn der Beobachtung zur Zeit b_0 und dem Eintritt der Zensierung, dem Zeitintervall also, ab dem die Untersuchungseinheit nicht mehr weiter verfolgt und beobachtet werden kann. Was man nicht weiß ist, wann das nächste Ereignis eingetreten und in welchen Zustand der Wechsel erfolgt wäre. Wir legen dabei den Zensierungsmechanismus des „random censoring“ zugrunde. Die Annahme besteht dann darin, daß die Zufallsvariable Zensierungszeit C_U für die Untersuchungseinheit U unabhängig ist von der Zufallsvariable Verweildauer beziehungsweise Eintrittszeit eines Ereignisses T_U . Beobachtbar ist jedoch nur $t = \min\{T_U, C_U\}$. Wir definieren dazu den Zensierungsindikator δ_U mit

$$\delta_U := \begin{cases} 1 & , \text{ falls } T_U \leq C_U , \text{ das heißt das Ereignis ist im } t\text{-ten} \\ & \text{Zeitintervall } [b_{t-1}, b_t) \text{ eingetreten.} \\ 0 & , \text{ falls } T_U > C_U , \text{ das heißt } t \text{ ist eine zensierte Zeit.} \end{cases} \quad (2.22)$$

Ferner nehmen wir wie auch Hamerle (1985b) und Fahrmeir und Tutz (1994, Abschnitt 9.2.3) an, daß der Zensierungsmechanismus „nichtinformativ“ ist. Das bedeutet, daß die Verteilung von C_U nicht von den für die Verteilung von T_U relevanten Variablen abhängt, im speziellen nicht von eventuell vorhandenen Regressionskoeffizienten. Ferner unterstellen wir, daß Rechtszensierung nur in der letzten Episode K_U , die eine Untersuchungseinheit erlebt, vorkommen kann und nicht in Zwischenepisoden k mit $1 \leq k \leq K_U - 1$.

Die Beschreibung und Darstellung weiterer Zensierungsmechanismen findet man beispielsweise in Kalbfleisch und Prentice (1980), Lawless (1982, Abschnitt 1.4), Hamerle und Tutz (1989) oder Lindsey (1993, Teilabschnitt 7.1.3).

2.5 Datenkodierung und Einbezug von Kovariablen

Nach der bisherigen Darstellung sind für jede Untersuchungseinheit U die Daten durch K_U und $(t_U^k, j_U^k, \delta_U^k)$, $k = 1, \dots, K_U$, gegeben. K_U bezeichnet die Anzahl der Episoden, die U erlebt, und $(t_U^k, j_U^k, \delta_U^k)$ sind die Daten zur k -ten Episode mit $k \leq K_U$. t_U^k ist dabei das Zeitintervall, in dem die k -te Episode endet, und j_U^k der Zustand, in den U im Zeitintervall t_U^k wechselt. Zusammen mit den Daten der $(k-1)$ -ten Episode ist die Verweildauer von U in der k -ten Episode durch $t_U^k - t_U^{k-1}$ gegeben. Weil wir davon ausgehen, daß nur die letzte Episode einer Untersuchungseinheit rechtszensiert sein kann, ist $\delta_U^k = 1 \quad \forall k = 1, \dots, K_U - 1$. $\delta_U^{K_U}$ setzen wir wie in (2.22) angegeben. Zur Formulierung unserer dynamischen Modelle in Kapitel 3, des Schätzansatzes in Kapitel 4 und der Schätzalgorithmen in Kapitel 5 ist es günstig, für die Daten eine andere, im mathematischen Sinne äquivalente Kodierung einzuführen, die einen natürlichen Zugang zur Verwendung von Zustandsraummodellen im Rahmen der Ereignisanalyse erleichtert. Dies geschieht in Anlehnung an Fahrmeir (1994) und Fahrmeir und Wagenpfeil (1995b), jetzt jedoch auf den Mehr-Episoden-Fall verallgemeinert.

Sei im folgenden $\bar{U} \in \mathbb{N}$ die Anzahl der Untersuchungseinheiten, die an einer Untersuchung teilgenommen haben und deren Daten vorliegen. Wir definieren für jede Untersuchungseinheit $U = 1, \dots, \bar{U}$, für jede Episode $k = 1, \dots, \bar{K} = \max_U \{K_U\}$ und für jedes Zeitintervall $t = 1, \dots, \bar{T}$ Risikoindikatoren r_{tU}^k durch

$$r_{tU}^k := \begin{cases} 1 & , \text{falls } U \text{ in der } k\text{-ten Episode ist und im Zeitintervall } t \\ & \text{noch "unter Risiko" steht.} \\ 0 & , \text{sonst.} \end{cases} \quad (2.23)$$

Wir nehmen dabei an, daß Zensierung, wenn überhaupt, am Ende eines Zeitintervalls auftritt. $r_{tU}^k = 1$ bedeutet, daß U zur Zeit t mindestens $(k-1)$ Zustandswechsel erlebt hat und weder der k -te Zustandswechsel bis zum Zeitintervall t noch eine Zensierung bis zum Zeitintervall $(t-1)$ eingetreten ist. Falls $k > K_U$, das heißt wenn k größer ist als die Episodenanzahl, die U erlebt, dann ist $r_{tU}^k = 0 \quad \forall t = 1, \dots, \bar{T}$. Ebenso gilt, falls der $(k-1)$ -te Zustandswechsel in t_{k-1} und der k -te Zustandswechsel in t_k stattfindet, daß $r_{tU}^k = 0 \quad \forall (t < t_{k-1} \vee t \geq t_k)$. Das heißt, wenn sich U im Zeitintervall t nicht in der k -ten Episode befindet, ist $r_{tU}^k = 0$. Am Beispiel der SOEP Arbeitslosendaten aus Teilabschnitt 1.1.5 bedeutet $r_{tU}^2 = 1$, daß die Person mit der Nummer U im Zeitintervall t zum zweiten Mal innerhalb des Beobachtungszeitraumes arbeitslos ist.

Die Risikomenge R_t^k definieren wir zu

$$R_t^k := \{U \mid r_{tU}^k = 1\}. \quad (2.24)$$

Damit ist R_t^k die Menge aller Untersuchungseinheiten, die mindestens k Episoden erlebt haben und im Zeitintervall t in der k -ten Episode noch „unter Risiko“ stehen. Die zuvor genannte Person U aus dem Arbeitslosendatensatz mit $r_{tU}^2 = 1$ ist also in R_t^2 enthalten. Weil wir angenommen haben, daß die Zensierung nur am Ende eines Zeitintervalls auftreten kann, enthält R_t^k auch all die Untersuchungseinheiten, die sich in der k -ten Episode befinden und im Intervall $[b_{t-1}, b_t)$ zensiert sind. Für den Fall der Zensierung am Anfang eines Zeitintervalls wären diese Untersuchungseinheiten auszuschließen.

Außerdem definieren wir Abgangsvektoren $y_{tU}^k \in \mathbb{R}^{\bar{J}}$ durch

$$y_{tU}^k := \begin{cases} e_j & , \text{falls } U \text{ in der } k\text{-ten Episode ist und im Zeitintervall } t \\ & \text{ein Zustandswechsel nach } j \text{ stattfindet.} \\ 0_{\bar{J}} & , \text{sonst.} \end{cases} \quad (2.25)$$

für $U = 1, \dots, \bar{U}$, $k = 1, \dots, \bar{K}$ und $t = 1, \dots, \bar{T}$. Wechselt beispielsweise die Person U aus dem Arbeitslosendatensatz, in dem $\bar{J}=9$ gilt, im Zeitintervall $t+1$ von der zweiten Arbeitslosigkeitsepisode in den Zustand mit der Nummer 5 (Rente oder Pension), dann ist $y_{t+1,U}^2 = e_5 \in \mathbb{R}^9$. Ist jedoch die zweite Arbeitslosigkeit für U eine zensierte Episode, dann gilt: $y_{tU}^2 = 0_9 \quad \forall t = 1, \dots, \bar{T}$. Tritt die Zensierung für die zweite Episode im Zeitintervall t auf, so ist $(y_{tU}^2 = y_{t+1,U}^2 = 0_9) \wedge (r_{tU}^2 = 1, r_{t+1,U}^2 = 0)$. Aus (2.25) ergibt sich insbesondere, daß innerhalb der k -ten Episode in einem festen Zeitintervall t nicht gleichzeitig in mehrere verschiedene Zustände gewechselt werden kann. Ferner gilt nach den Definitionen (2.24) und (2.25):

$$(U \in \mathbb{R}_t^k) \Rightarrow (y_{1U}^l = y_{2U}^l = \dots = y_{t-1,U}^l = 0_{\bar{J}} \quad \forall l \in \{k, \dots, \bar{K}\}).$$

Das heißt, wenn sich die Untersuchungseinheit U im Zeitintervall t noch in der Risikomenge der k -ten Episode befindet, dann kann U zuvor keinen Zustandswechsel von der l -ten in die $(l+1)$ -te Episode vollzogen haben für alle $l \in \{k, \dots, \bar{K} - 1\}$.

Zur Vereinfachung der Notation fassen wir in den Vektoren

$$\begin{aligned} r_t^k &:= (r_{tU}^k, U = 1, \dots, \bar{U}), \\ y_t^k &:= (y_{tU}^k, U = 1, \dots, \bar{U}), \\ r_t^{k*} &:= (r_t^l, l = 1, \dots, k) \in \mathbb{R}^{\bar{U} \cdot k}, \\ y_t^{k*} &:= (y_t^l, l = 1, \dots, k) \in \mathbb{R}^{\bar{U} \cdot k \cdot \bar{J}} \end{aligned} \quad (2.26)$$

die Risikoindikatoren beziehungsweise Abgangsvektoren aller Untersuchungseinheiten zur Zeit t für die k -te respektive bis zur k -ten Episode zusammen. Den Verlauf dieser Kenngrößen bis zum Zeitintervall t bezeichnen wir mit

$$y_{t*}^{k*} := (y_1^{k*}, \dots, y_t^{k*}), \quad r_{t*}^{k*} := (r_1^{k*}, \dots, r_t^{k*}). \quad (2.27)$$

Für den Ein-Episoden-Fall ($\bar{K} = 1$) setzen wir

$$\begin{aligned} \mathbf{r}_t &:= \mathbf{r}_t^k = \mathbf{r}_t^{k*}, \mathbf{y}_t := \mathbf{y}_t^k = \mathbf{y}_t^{k*}, \\ \mathbf{r}_t^* &:= \mathbf{r}_{t^*}^{k*} = (\mathbf{r}_1, \dots, \mathbf{r}_t), \mathbf{y}_t^* := \mathbf{y}_{t^*}^{k*} = (\mathbf{y}_1, \dots, \mathbf{y}_t). \end{aligned} \quad (2.28)$$

Die Definitionen von \mathbf{r}_{tU}^k und \mathbf{y}_{tU}^k wurden so gewählt, daß bei Kenntnis von \mathbf{K}_U und $(\mathbf{t}_U^k, \mathbf{j}_U^k, \delta_U^k)$, $k = 1, \dots, \mathbf{K}_U$, die Werte von \mathbf{r}_{tU}^k und \mathbf{y}_{tU}^k festgelegt sind und umgekehrt. Es sei im folgenden $\lambda_j^k(t)$ die übergangsspezifische Hazardrate der k -ten Episode für die Untersuchungseinheit U . Dann gilt

$$\begin{aligned} \lambda_j^k(t) &\stackrel{(2.13)}{=} \text{pr}(\mathbf{T}_k = \mathbf{t}, \mathbf{J}_k = \mathbf{j} | \mathbf{T}_k \geq \mathbf{t}, \mathbf{H}_{k-1}) = \\ &= \text{pr}(\mathbf{y}_{tU}^k = \mathbf{e}_j | \mathbf{r}_{t^*}^{k*}, \mathbf{y}_{(t-1)^*}^{k*}, \mathbf{y}_t^{(k-1)*}, \mathbf{y}_{1U}' = \mathbf{y}_{2U}' = \dots = \mathbf{y}_{t-1,U}' = \mathbf{0}_j \quad \forall l \in \{k, \dots, \bar{K}\}). \end{aligned} \quad (2.29)$$

In (2.29) wurde aus Notationsgründen nicht zwischen einer Zufallsvariable und dem Wert einer Zufallsvariable unterschieden.

Zusätzlich enthalten viele Ereignisdaten Kovariable \mathbf{x}_{tU} , die von jeder Untersuchungseinheit U vorliegen und möglicherweise über die Zeit t variieren. Wir wollen sie in dem Vektor

$$\mathbf{x}_t := (\mathbf{x}_{tU}, U = 1, \dots, \bar{U}) \quad (2.30)$$

zusammenfassen. Aus technischen Gründen definieren wir noch für den Mehr-Episoden-Fall $\mathbf{x}_t^k := \mathbf{x}_t$, $k = 1, \dots, \bar{K}$, das heißt wir führen weitere Größen ein, die für festes t mit dem Wert \mathbf{x}_t vollständig korreliert sind. Das ist eine Maßnahme, um in Abschnitt 4.2 die penalisierte Loglikelihood in Produktform über die Episoden k schreiben zu können, wie es die gewöhnliche Likelihood in Hamerle (1985b) nahelegt. Die zeitliche Entwicklung bis zum Zeitintervall t schreiben wir in den Vektor

$$\mathbf{x}_t^* := (\mathbf{x}_1, \dots, \mathbf{x}_t), \quad (2.31)$$

und für den Mehr-Episoden-Fall in $\mathbf{x}_{t^*}^{k*} := (\mathbf{x}_1^{k*}, \dots, \mathbf{x}_t^{k*})$ mit $\mathbf{x}_t^{k*} := (\mathbf{x}_t^l, l = 1, \dots, k)$. Für den Ein-Episoden-Fall gilt: $\mathbf{x}_{t^*}^{k*} = \mathbf{x}_t^*$. Weil manche Untersuchungseinheiten früher oder später als andere aus der Beobachtung ausscheiden, können im allgemeinen zur Zeit t nicht für alle U Kovariablenwerte erhoben werden. Wann eine solche t - U -Kombination vorliegt, geht aus \mathbf{r}_{tU}^k und \mathbf{y}_{tU}^k mit $U = 1, \dots, \bar{U}$, $k = 1, \dots, \bar{K}$ und $t = 1, \dots, \bar{T}$ hervor. Es wird dann aus technischen Gründen $\mathbf{x}_{tU} = \mathbf{0}$ gesetzt. Dies hat keinen Einfluß auf die Schätzung, denn derartige t - U -Kombinationen kommen im Schätzansatz von Abschnitt 4.2 nicht vor.

Die Kovariablen können deterministisch oder stochastisch sein. Stochastische Kovariable werden in der Bedingung der Hazardrate aufgenommen:

$$\begin{aligned}\lambda_j^k(t|\mathbf{x}_t^{k*}) &= \text{pr}(T_k = t, J_k = j | T_k \geq t, H_{k-1}, \mathbf{x}_t^{k*}) = \\ &= \text{pr}(\mathbf{y}_{tU}^k = \mathbf{e}_j | \mathbf{r}_t^{k*}, \mathbf{y}_{(t-1)*}^{k*}, \mathbf{y}_t^{(k-1)*}, \mathbf{x}_t^{k*}, \mathbf{y}_{1U}^l = \mathbf{y}_{2U}^l = \dots = \mathbf{y}_{t-1,U}^l = \mathbf{0}_j \ \forall l \in \{k, \dots, \bar{K}\}).\end{aligned}\tag{2.32}$$

Für feste, nicht über die Zeit variierende Kovariable setzen wir $\mathbf{x}_U := \mathbf{x}_{1U} = \dots = \mathbf{x}_{\bar{T},U}$.

Kalbfleisch und Prentice (1980) unterscheiden ferner bei zeitabhängigen Einflußgrößen zwischen internen und externen Kovariablen. Was in diesem Zusammenhang bei der Schätzung zu beachten ist, wird in den Abschnitten 3.2, 4.1 und 4.2 besprochen.

Das Konzept der Hazardrate besteht darin, einen Regressionsansatz für die Hazardrate λ zu wählen, der den Baseline-Effekt und möglicherweise weitere Kovariable enthält und deren Einfluß auf λ funktional festlegt. Unter der Voraussetzung, daß die Hazardraten für jede Episode von verschiedenen Regressionskoeffizienten abhängen, kann man die Schätzung im Mehr-Episoden-Fall getrennt für jede Episode k durchführen (siehe dazu Hamerle, 1985b, Abschnitt 3). Deshalb ist im nächsten Kapitel die Darstellung der dynamischen Modelle zur Analyse von Ereignisdaten auf eine Episode k konzentriert.

3. Dynamische Modelle

Diskrete Ereignisdaten wie auch diskrete Zeitreihendaten haben unter anderem zwei wesentliche Strukturmerkmale: Die Daten variieren über die Zeit und sie sind in diskreten Zeitintervallen erhoben. Für eine solche Datenlage eignen sich insbesondere Zustandsraummodelle, weil auch ihnen die Merkmale „diskret und dynamisch“ zueigen sind.

Im folgenden werden wir zunächst für Zeitreihenbeobachtungen y_t das lineare Zustandsraummodell und dessen Erweiterung auf die Klasse der Exponentialfamilienverteilungen, das verallgemeinerte Zustandsraummodell, darstellen. In Abschnitt 3.1 wollen wir zeigen, wie das verallgemeinerte Zustandsraummodell für die Analyse von Ereignisdaten verwendet werden kann. Damit kommen wir zum multinomialen Zustandsraummodell. In Abschnitt 3.2 beschreiben wir, wie das piecewise exponential Modell, das Gamerman (1991) zur Untersuchung von Survival-Daten einsetzt, in Zustandsraumform gebracht werden kann. Damit bilden Zustandsraummodelle ein Rahmenmodell für die Ereignisanalyse und eine Grundlage für die Schätzung, die in Kapitel 4 besprochen wird.

Lineare Zustandsraummodelle mit Normalverteilung wurden bereits in der Raumfahrt und Technik bei Zeitreihendaten ($\bar{U} = 1$) erfolgreich eingesetzt. Die Beobachtungen y_t , $t = 1, \dots, \bar{T}$, waren zum Beispiel die jeweiligen Raketenpositionen auf dem Radarschirm zu den Zeiten t . Grundlegend hierfür ist das Buch von Anderson und Moore (1979). Diese Modelle können als dynamische Verallgemeinerung des klassischen linearen Modells betrachtet werden und sie bestehen aus zwei Teilen: Im

- Beobachtungsmodell

$$y_t = Z_t \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \Sigma_t), \quad t = 1, \dots, \bar{T}, \quad (3.1)$$

wird angenommen, daß die Zeitreihenbeobachtungen $y_t, t = 1, \dots, \bar{T}$, beschrieben werden können durch einen systematischen Teil $Z_t \alpha_t$, zu dem ein zufälliger Teil ε_t addiert wird. Für die Sequenz $\{\varepsilon_t\}$ der Zufallsvariablen ε_t wird „weißes Rauschen“ angenommen, das heißt die ε_t sollen paarweise unkorreliert sein. Z_t ist die Design-Matrix, die auch zeitvariierende Kovariable x_t enthalten kann, und der Zustandsvektor $\alpha_t \in \mathbb{R}^n$, der die Kovariableneffekte enthält, gewichtet den Einfluß der entsprechenden Komponenten von Z_t auf y_t . Die Vektoren $\alpha_t, t = 1, \dots, \bar{T}$, sind die nicht beobachtbaren, zu schätzenden

Zustände des Zustandsraummodells. Weil man in (3.1) im Zeitreihenfall für jeden Vektor α_t nur eine Beobachtung y_t zur Verfügung hat, führt eine gewöhnliche Schätzung von α_t zu unbrauchbaren Ergebnissen. Deshalb werden an die Sequenz $\{\alpha_t\}$ der Zustände α_t , $t = 1, \dots, \bar{T}$, weitere Restriktionen gestellt. Wir nehmen dabei den Bayes'schen Standpunkt ein und formulieren das lineare

- Übergangsmodell

$$\alpha_t = F_t \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(0_n, Q_t), \quad t = 1, \dots, \bar{T}, \quad \alpha_0 \sim N(a_0, Q_0). \quad (3.2)$$

$F_t \in \mathbb{R}^{n,n}$ wird als Übergangsmatrix bezeichnet und für die Sequenz $\{\xi_t\}$ der n-dimensionalen Zufallsvariablen ξ_t unterstellen wir wieder „weißes Rauschen“.

Außerdem sollen die Zufallsvariablen $\{\varepsilon_t\}$, $\{\xi_t\}$ und α_0 auch untereinander unkorreliert sein, das heißt

$$\begin{aligned} E(\varepsilon_s \xi_t') &= 0_{n,n} \quad \forall s, t = 1, \dots, \bar{T} \quad \text{und} \\ E(\varepsilon_t \alpha_0') &= 0_{n,n}, \quad E(\xi_t \alpha_0') = 0_{n,n} \quad \forall t = 1, \dots, \bar{T}. \end{aligned} \quad (3.3)$$

Wegen der Normalverteilungsannahme in (3.1) und (3.2) ist die Unkorreliertheit in (3.3) äquivalent mit der Unabhängigkeit von $\{\varepsilon_t\}$, $\{\xi_t\}$ und α_0 . Die Größen Σ_t, Q_t, Q_0 und a_0 werden als Hyperparameter bezeichnet. Q_0 und a_0 legen die Start-Priori-Verteilung des n-dimensionalen Zufallsvektors α_0 fest.

Das lineare Zustandsraummodell (3.1), (3.2) enthält viele strukturelle Zeitreihenmodelle als Spezialfälle. Für eine umfassende Darstellung siehe etwa Harvey (1992). Viele Trendmodelle wie etwa den „random walk erster Ordnung“ RW1 oder „zweiter Ordnung“ RW2 oder das lokal lineare Trendmodell und auch verschiedene Saisonmodelle, die man zur Beschreibung des Übergangs der Komponenten von α_{t-1} nach α_t verwenden kann, lassen sich in die allgemeine Form des Übergangsmodells (3.2) bringen (vergleiche dazu Fahrmeir und Tutz, 1994, Teilabschnitt 8.1.1).

Sei beispielsweise $\dim(\alpha_t) = 1$ und der Übergang von α_{t-1} nach α_t soll mit einem RW1 modelliert werden. Dann ist $F_t = Z_t = 1$ und

$$\alpha_t = \alpha_{t-1} + \xi_t, \quad y_t = \alpha_t + \varepsilon_t. \quad (3.4)$$

Wenn aber der Übergang von τ_{t-2} und τ_{t-1} nach τ_t mit einem RW2 beschrieben werden soll, das heißt $\tau_t = 2\tau_{t-1} - \tau_{t-2} + \mathbf{v}_t$, $\mathbf{v}_t \sim N(0, \mathbf{Q}_t)$, und $\dim(\tau_t) = 1$, dann setzen wir mit $\alpha_t = (\tau_t, \tau_{t-1})'$

$$\alpha_t = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix} \cdot \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(0_2, \mathbf{Q}_t), \quad \mathbf{Q}_t = \begin{pmatrix} \mathbf{q}_t & 0 \\ 0 & 0 \end{pmatrix}, \quad \mathbf{q}_t \geq 0,$$

$$\mathbf{y}_t = (1, 0) \cdot \alpha_t + \varepsilon_t. \quad (3.5)$$

(3.5) ist die Zustandsraumform für das RW2-Modell (fast sicher, weil in \mathbf{Q}_t die Varianz der zweiten Komponente von ξ_t gleich Null ist).

Mit der Wahl von $\mathbf{F}_t = \mathbf{I}_n$, $\mathbf{Q}_0 = \mathbf{Q}_t = \mathbf{0}_{n,n}$, $t = 1, \dots, \bar{T}$, im linearen Zustandsraummodell erhält man (fast sicher) das klassische lineare Modell und damit den Fall von fixen, zeitkonstanten Parametern. Vergleicht man nun das lineare Zustandsraummodell (3.1), (3.2) mit dem klassischen linearen Modell, so fallen zwei wesentliche Unterschiede auf: Zum einen sind die Regressionskoeffizienten α_t dynamisch, können also über die Zeit variieren, und heißen jetzt Zustände, zum anderen sind sie keine festen, zu schätzenden Parameter mehr sondern Zufallsvariable. Eine Gemeinsamkeit von beiden Modellen ist die Annahme der Normalverteilung.

Für fest vorgegebenes $k \in \{1, \dots, \bar{K}\}$ und $U \in \{1, \dots, \bar{U}\}$ sind jedoch die Zeitreihenbeobachtungen $\{\mathbf{y}_{tU}^k\}$, $t = 1, \dots, \bar{T}$, die bei Ereignisdatensätzen vorliegen, Sequenzen von Ja/Nein-Entscheidungen darüber, ob die Untersuchungseinheit U , wenn sie sich in der k -ten Episode befindet, im Zeitintervall t in den Zustand j wechselt oder nicht, gegeben die Vorgeschichte. Die j -te Komponente y_{tUj}^k von \mathbf{y}_{tU}^k kann man daher als Ergebnis eines Bernoulli-Experiments auffassen, wobei für die Eintrittswahrscheinlichkeiten gelten muß:

$$\text{pr}(\mathbf{y}_{tU}^k = \mathbf{0}_j) + \sum_{j=1}^J \text{pr}(\mathbf{y}_{tUj}^k = 1) = 1.$$

Zur Modellierung dieser Daten ist wie im Falle von kategorialen Zeitreihen die Multinomialverteilung geeignet. Die Normalverteilungsannahme in (3.1), (3.2) ist daher für Ereignisdaten zu restriktiv. Aus diesem Grunde betrachten wir eine Erweiterung des linearen Zustandsraummodells, bei der das lineare Beobachtungsmodell mit Normalverteilung für die Elemente \mathbf{y}_t der Zeitreihe $\{\mathbf{y}_t\}$, $t = 1, \dots, \bar{T}$, ersetzt wird durch die

Annahme, daß die bedingte Verteilung von y_t , gegeben die Vorgeschichte α_t , $y_{t-1}^* := (y_1, \dots, y_{t-1})$ und $x_t^* := (x_1, \dots, x_t)$, zur Klasse der einfachen Exponentialfamilie mit Gewichtsparameter ω_{U} gehört. Für eine Darstellung der entsprechenden Dichteformel siehe Fahrmeir und Tutz (1994, Formeln (2.1.4) und (3.1.7)). Die Multinomial- und die Poissonverteilung gehören der einfachen Exponentialfamilie an (siehe etwa Fahrmeir und Hamerle, 1984, Kapitel 2, Abschnitt 5.2).

Damit gelangen wir zum verallgemeinerten Zustandsraummodell, das wie das lineare Zustandsraummodell aus zwei Stufen besteht:

- Beobachtungsmodell

$$y_t | y_{t-1}^*, x_t^*, \alpha_t \sim p(y_t | y_{t-1}^*, x_t^*, \alpha_t) :=$$

= Dichte der einfachen Exponentialfamilie mit Gewichtsparameter ω_{U} ,

wobei wie im verallgemeinerten linearen Modell

$$E(y_t | y_{t-1}^*, x_t^*, \alpha_t) =: \mu_t(\alpha_t) = h(\eta_t), \quad \eta_t := Z_t \alpha_t, \quad t = 1, \dots, \bar{T}. \quad (3.6)$$

p bezeichnet dabei, wie auch im folgenden, die Dichte der durch die jeweiligen Zufallsvariablen induzierten Verteilung. $h: \mathbb{R}^{\bar{J}} \rightarrow \mathbb{R}^{\bar{J}}$, $\eta_t \rightarrow h(\eta_t)$, ist eine 2-mal stetig differenzierbare Responsefunktion, $\eta_t \in \mathbb{R}^{\bar{J}}$ der lineare Prädiktor, $Z_t \in \mathbb{R}^{\bar{J}, n}$ und $\alpha_t \in \mathbb{R}^n$ wie in (3.1) die Design-Matrix beziehungsweise die zu schätzenden Zustände. Eine Eigenschaft der Exponentialfamilie besteht darin, daß durch den Regressionsansatz in (3.6) auch gleichzeitig die Varianz $\Sigma_t(\alpha_t) := \text{var}(y_t | y_{t-1}^*, x_t^*, \alpha_t)$ mitspezifiziert wird (siehe dazu auch Fahrmeir und Tutz, 1994, Teilabschnitt 2.1.2, und für die Möglichkeiten zur Wahl von h die Teilabschnitte 2.1.3 - 2.1.5 und die Abschnitte 3.2 und 3.3).

Für den Übergang von α_{t-1} nach α_t wird das lineare Übergangsmodell mit Normalverteilung beibehalten:

- Übergangsmodell

$$\alpha_t | \alpha_{t-1} \sim p(\alpha_t | \alpha_{t-1}) = N(F_t \alpha_{t-1}, Q_t), \quad t = 1, \dots, \bar{T}, \quad \alpha_0 \sim p(\alpha_0) = N(a_0, Q_0). \quad (3.7)$$

$p(\cdot) = N(\cdot)$ in (3.7) ist symbolisch gemeint und bedeutet, daß $p(\cdot)$ die (bedingte) Dichte der Normalverteilung bezeichnet.

Es sei im folgenden $\alpha_t^* := (\alpha'_0, \alpha'_1, \dots, \alpha'_t)'$. Für eine vollständige Modellspezifikation treffen wir noch die nachstehenden drei Vereinbarungen:

$$p(y_t | y_{t-1}^*, x_t^*, \alpha_t^*) = p(y_t | y_{t-1}^*, x_t^*, \alpha_t), \quad t = 1, \dots, \bar{T}. \quad (3.8)$$

Wir nehmen also an, daß die Beobachtung y_t , gegeben y_{t-1}^*, x_t^* , nur vom gegenwärtigen Zustand α_t und nicht von den vergangenen Zuständen α_{t-1}^* abhängt. Diese bedingte Unabhängigkeitsannahme wird implizit auch im linearen Zustandsraummodell mit (3.3) gemacht.

Für den Fall von stochastischen Kovariablen x_t setzen wir ferner voraus, daß

$$p(x_t | y_{t-1}^*, x_{t-1}^*, \alpha_{t-1}^*) = p(x_t | y_{t-1}^*, x_{t-1}^*), \quad t = 1, \dots, \bar{T}, \quad (3.9)$$

das heißt der Kovariablenprozeß $\{x_t\}$ gegeben die Vergangenheitsinformation y_{t-1}^*, x_{t-1}^* , ist unabhängig von den vergangenen Zuständen α_{t-1}^* . (3.9) wird im Fall von deterministischen Kovariablen nicht benötigt.

Als dritten Punkt nehmen wir an, daß der Zustandsprozeß $\{\alpha_t\}$ Markoveigenschaft besitzt, genauer:

$$p(\alpha_t | y_{t-1}^*, x_t^*, \alpha_{t-1}^*) = p(\alpha_t | \alpha_{t-1}), \quad t = 1, \dots, \bar{T}. \quad (3.10)$$

Die Forderung (3.10) ist implizit im Übergangsmodell (3.2) beziehungsweise (3.7) zusammen mit (3.3) enthalten.

Das Beobachtungsmodell (3.6) hat die Form eines verallgemeinerten linearen Modells, jedoch mit zeitvariierenden Regressionskoeffizienten. Deshalb wird das verallgemeinerte Zustandsraummodell (3.6), (3.7) auch als dynamisches, verallgemeinertes lineares Modell bezeichnet. Wählt man in (3.6) für $p(y_t | y_{t-1}^*, x_t^*, \alpha_t)$ die Dichte der Normalverteilung, die ebenfalls in der Klasse der Exponentialfamilienverteilungen enthalten ist, und für die Responsefunktion $h: \mathbb{R}^{\bar{J}} \rightarrow \mathbb{R}^{\bar{J}}$ die identische Abbildung, $\mu_t(\alpha_t) = h(\eta_t) = \eta_t = Z_t \alpha_t$, so ergibt sich als Spezialfall von (3.6), (3.7) das lineare Zustandsraummodell (3.1), (3.2).

3.1 Multinomiales Zustandsraummodell

Wir betrachten, wie im vorherigen Abschnitt bereits angedeutet, die Beobachtungen $y_{tU}^k \in \mathbb{R}^{\bar{J}}$ für eine feste Episode k und für eine gegebene Untersuchungseinheit U als kategoriale Zeitreihe mit $t=1, \dots, \bar{T}$. Für die Schätzung in den Abschnitten 4.1 und 4.2 sind jedoch nur die Zeitintervalle t relevant, für die gilt: $t \in T_{kU}^* := \{t | U \in R_t^k\}$. Befindet sich U nämlich nicht in der k -ten Episode, also $U \notin R_t^k$, so kann auch kein Zustandswechsel aus der k -ten Episode heraus stattfinden. Die bedingte Eintrittswahrscheinlichkeit der j -ten Kategorie ist für $t \in T_{kU}^*$ gegeben durch

$$\begin{aligned} \pi_j^k &:= \text{pr}(y_{tU}^k = e_j | r_{t^*}^{k*}, y_{(t-1)^*}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, \alpha_t^k) \stackrel{\text{da } U \in R_t^k}{=} \\ &= \text{pr}(y_{tU}^k = e_j | r_{t^*}^{k*}, y_{(t-1)^*}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, \alpha_t^k, y_{1,U}^l = y_{2,U}^l = \dots = y_{t-1,U}^l = 0_j \quad \forall l \in \{k, \dots, \bar{K}\}) = \\ &\stackrel{(2.32)}{=} \lambda_j^k(t | x_{t^*}^{k*}, \alpha_t^k), \quad j = 1, \dots, \bar{J}. \end{aligned} \quad (3.11)$$

Für $t \notin T_{kU}^*$ sei dann $\pi_j^k := \text{pr}(y_{tU}^k = e_j | r_{t^*}^{k*}, y_{(t-1)^*}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*})$ unabhängig von α_t^k . Diese Annahme werden wir in Abschnitt 4.1 rechtfertigen. Der Wert von π_j^k ist identisch mit der j -ten Komponente des Erwartungswertvektors $E(y_{tU}^k | r_{t^*}^{k*}, y_{(t-1)^*}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, \alpha_t^k)$, die wir mit $\mu_{tU}^k(\alpha_t^k)$ bezeichnen wollen. In (3.11) gehen wir davon aus, daß die Hazardraten $\lambda_j^k(t | x_{t^*}^{k*}, \alpha_t^k)$ für jede Episode $k \in \{1, \dots, \bar{K}\}$ von verschiedenen Regressionskoeffizienten beziehungsweise Zustandsvektoren α_t^k abhängen. Die Zustände $\alpha_t^k \in \mathbb{R}^n$ variieren somit über die Zeit t und die Episode k . α_t^l mit $l \neq k$, $t = 1, \dots, \bar{T}$, soll dabei keinen Einfluß auf $\lambda_j^k(t | x_{t^*}^{k*})$ haben. Insgesamt setzen wir das Beobachtungsmodell wie für kategoriale Zeitreihen:

- Beobachtungsmodell

$$\begin{aligned} y_{tU}^k | r_{t^*}^{k*}, y_{(t-1)^*}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, \alpha_t^k &\sim p(y_{tU}^k | r_{t^*}^{k*}, y_{(t-1)^*}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, \alpha_t^k) = M(1, \pi_1^k, \dots, \pi_J^k). \\ \mu_{tU}^k(\alpha_t^k) = \pi_j^k &= h_j(\eta_{tU}^k), \quad \mathbb{R}^{\bar{J}} \ni \eta_{tU}^k := Z_{tU}^k \cdot \alpha_t^k, \quad t = 1, \dots, \bar{T}, \quad j = 1, \dots, \bar{J}. \end{aligned} \quad (3.12)$$

Das lineare Übergangsmodell mit Normalverteilung aus (3.7) wird beibehalten:

- Übergangsmodell

$$\alpha_t^k | \alpha_{t-1}^k \sim p(\alpha_t^k | \alpha_{t-1}^k) = N(F_t^k \alpha_{t-1}^k, Q_t^k), \quad t = 1, \dots, \bar{T},$$

$$\alpha_0^k \sim p(\alpha_0^k) = N(a_0^k, Q_0^k). \quad (3.13)$$

(3.12) zusammen mit (3.13) bezeichnen wir als multinomiales Zustandsraummodell der k -ten Episode. Die Design-Matrix Z_{tU}^k kann sowohl zeitvariierende und episodenspezifische Kovariable enthalten, als auch exogene Variable aus anderen Episoden der Untersuchungseinheit U , beispielsweise die Verweildauer von U in der $(k-1)$ -ten Episode. Für die Übergangsmatrizen F_t^k und Fehler-Kovarianz-Matrizen Q_t^k aus dem Übergangsmodell nehmen wir an, daß sie über die Zeit und über die Episode variieren können. η_{tU}^k ist der lineare Prädiktor und $h_j: \mathbb{R}^{\bar{J}} \rightarrow \mathbb{R}$ in (3.12) ist die j -te Komponente der Responsefunktion $h: \mathbb{R}^{\bar{J}} \rightarrow \mathbb{R}^{\bar{J}}$, $h(\eta_{tU}^k) := (h_1(\eta_{tU}^k), \dots, h_{\bar{J}}(\eta_{tU}^k))'$. Sie modelliert in (3.12) die Eintrittswahr-

scheinlichkeit und damit nach (3.11) die abgangsspezifische Hazardrate. Da $0 \leq \sum_{j=1}^{\bar{J}} \pi_j^k \leq 1$,

kommen für die Wahl von h_j insbesondere Funktionen in Frage, für die gilt:

$\sum_{j=1}^{\bar{J}} h_j(\eta_{tU}^k) \in (0, 1)$. Sei η_{tUj}^k der lineare Prädiktor für den Abgang in den Zustand j und

damit $\eta_{tU}^k = (\eta_{tU,1}^k, \dots, \eta_{tU,\bar{J}}^k)'$. Dann ist beim dynamischen multinomialen Logit-Modell die Responsefunktion j gegeben durch

$$h_j(\eta_{tU}^k) := \frac{\exp(\eta_{tUj}^k)}{1 + \sum_{i=1}^{\bar{J}} \exp(\eta_{tUi}^k)}, \quad j = 1, \dots, \bar{J}. \quad (3.14)$$

h in (3.14) ist die inverse Funktion der natürlichen Linkfunktion zur Multinomialverteilung. Der Beweis findet sich in Anhang B. Wir bezeichnen diese Funktion h deshalb als natürliche Responsefunktion der Multinomialverteilung. Welche algorithmischen und numerischen Vorteile sich bei der Wahl von h in (3.14) ergeben, wird in Abschnitt 5.4 behandelt.

Weitere Möglichkeiten für die Wahl von h bestehen darin, Responsefunktionen aus zeitstetigen Modellen abzuleiten. Ausgehend von einem Proportional-Hazards-Modell, bei dem

die Baseline-Hazardrate nicht vom Abgangszustand j abhängt, leiten Hamerle und Tutz (1989, Abschnitt 6.1) das folgende Modell her:

$$\lambda_j^k(t|\mathbf{x}_t^{k*}, \alpha_t^k) = h_j(\eta_{tU}^k) := \frac{\exp(\eta_{tUj}^k)}{\sum_{i=1}^{\bar{J}} \exp(\eta_{tUi}^k)} \{1 - \exp(-\sum_{i=1}^{\bar{J}} \exp(\eta_{tUi}^k))\}, \quad j = 1, \dots, \bar{J}. \quad (3.15a)$$

Für den Ein-Zustands-Fall $\bar{J} = 1$ reduziert sich (3.15a) zum dynamischen gruppierten Cox-Modell

$$\lambda^k(t|\mathbf{x}_t^{k*}, \alpha_t^k) = h(\eta_{tU}^k) = 1 - \exp\{-\exp(\eta_{tU}^k)\}. \quad (3.15b)$$

Da wir in (3.15) zeitabhängige Regressionskoeffizienten α_t^k , $t = 1, \dots, \bar{T}$, verwenden, verliert $\lambda_j^k(t|\mathbf{x}_t^{k*}, \alpha_t^k)$ in (3.15) die Proportional-Hazards-Eigenschaft.

Die Verwendung von ordinalen Modellen mit sequentieller oder kumulativer Struktur wäre, bei entsprechender Datenlage, ebenso möglich (vergleiche hierzu Fahrmeir und Tutz, 1994, Kapitel 3 und Teilabschnitt 9.3.1).

Zu einer vollständigen Modellspezifikation gehört noch die Wahl der Design Matrix \mathbf{Z}_{tU}^k in (3.12). Es sei im folgenden τ_{tj}^k der Baseline-Effekt für den Abgangszustand j , $j \in \{1, \dots, \bar{J}\}$. x_u sei ohne Einschränkung der Allgemeinheit eine zeitunabhängige Kovariable, beispielsweise das Geschlecht einer unter Beobachtung stehenden arbeitslosen Person aus dem SOEP Arbeitslosendatensatz in Teilabschnitt 1.1.5, und β_t^k der zugehörige Kovariableneffekt im Zeitintervall t . Für den Parameterübergang im Übergangsmodell wählen wir den RW1. Dann wäre der lineare Prädiktor gegeben durch $\eta_{tUj}^k = \tau_{tj}^k + \mathbf{x}_u \cdot \beta_t^k$, $j = 1, \dots, \bar{J}$. Für die Zustände setzen wir $\alpha_t^k = (\tau_{t,1}^k, \tau_{t,2}^k, \dots, \tau_{t,\bar{J}}^k, \beta_t^k)'$ und für die Design-Matrix

$$\mathbf{Z}_{tU}^k = \begin{pmatrix} 1 & 0 & \mathbf{x}_u \\ & \ddots & \vdots \\ 0 & 1 & \mathbf{x}_u \end{pmatrix} \in \mathbb{R}^{\bar{J}, \bar{J}+1}. \quad (3.16)$$

Weil β_t^k nicht von der Abgangskategorie j abhängt, wird ein solcher Effekt als global bezeichnet. Sollen die Kovariableneffekte auch abgangsspezifisch sein, das heißt $\eta_{tUj}^k = \tau_{tj}^k + \mathbf{x}_u \beta_{tj}^k$, $j = 1, \dots, \bar{J}$, dann setzen wir in (3.12)

$$\mathbf{Z}_{tU}^k = \begin{pmatrix} 1 & & 0 & \mathbf{x}_u & & 0 \\ & \ddots & & & \ddots & \\ 0 & & 1 & 0 & & \mathbf{x}_u \end{pmatrix} \in \mathbb{R}^{\bar{J}, 2 \cdot \bar{J}} \quad (3.17)$$

und $\alpha_t^k = (\tau_{t,1}^k, \tau_{t,2}^k, \dots, \tau_{t,\bar{J}}^k, \beta_{t,1}^k, \beta_{t,2}^k, \dots, \beta_{t,\bar{J}}^k)'$.

Die Vergangenheitsinformation aus der gegenwärtigen oder den vergangenen Episoden, $\mathbf{y}_{(t-1)^*}^{k*}$, $\mathbf{y}_t^{(k-1)*}$, \mathbf{x}_t^{k*} , Saisondummies und zeitabhängige Kovariable \mathbf{x}_t wie zum Beispiel die Zeitintervalllänge $\mathbf{w}_t := \mathbf{b}_t - \mathbf{b}_{t-1} = \mathbf{x}_t$, wobei $\mathbf{b}_0 \leq \mathbf{b}_1 \leq \mathbf{b}_2 \leq \dots \leq \mathbf{b}_{\bar{T}}$ die Zeitintervallgrenzen wie in Abschnitt 2.1 sind, können ebenfalls in die Design-Matrix aufgenommen werden. Für eine ausführliche Behandlung sei auf Fahrmeir und Tutz (1994, Kapitel 3, 8 und 9) verwiesen. Beispiele finden sich auch in Kapitel 6 bei der Modellierung der Datensätze aus Abschnitt 1.1.

3.2 Piecewise exponential Modell

Wir betrachten im folgenden den Ein-Episoden-Ein-Zustands-Fall, also $\bar{J} = \bar{K} = 1$. Friedman (1982) benutzt in diesem Zusammenhang das piecewise exponential Modell, um mit Hilfe des Maximum-Likelihood-Verfahrens simultan die Hazardrate und Kovariablen-effekte zu schätzen. Gamerman (1991) verallgemeinert dieses Vorgehen auf den Fall von zeitvariierenden Effekten. Es sei die Zeitachse wie in Abschnitt 2.1 gruppiert in $\bar{T}+1$ Zeitintervalle $[\mathbf{b}_0, \mathbf{b}_1), [\mathbf{b}_1, \mathbf{b}_2), \dots, [\mathbf{b}_{\bar{T}-1}, \mathbf{b}_{\bar{T}}), [\mathbf{b}_{\bar{T}}, \infty)$ mit $\mathbf{0} = \mathbf{b}_0 \leq \mathbf{b}_1 \leq \mathbf{b}_2 \leq \dots \leq \mathbf{b}_{\bar{T}}$. Die Zeitintervalllängen $\mathbf{w}_t := \mathbf{b}_t - \mathbf{b}_{t-1}$, $t = 1, \dots, \bar{T}$, können dabei von Intervall zu Intervall unterschiedlich sein. Eine stetige Zufallsvariable T ist „piecewise exponential“-verteilt, in Zeichen $T \sim \text{PE}(\lambda, \mathbf{b})$, wenn für ihre Hazardrate gilt:

$$\lambda(t) = \begin{cases} \lambda_1 & , \text{ falls } t \in [\mathbf{b}_0, \mathbf{b}_1) \\ \lambda_t & , \text{ falls } t \in [\mathbf{b}_{t-1}, \mathbf{b}_t) \text{ mit } t \in \{2, \dots, \bar{T}\} \\ \lambda_{\bar{T}+1} & , \text{ falls } t \in [\mathbf{b}_{\bar{T}}, \infty) \end{cases} \quad (3.18)$$

mit $\lambda := (\lambda_1, \dots, \lambda_{\bar{T}+1})'$ und dem Zeitgrid $\mathbf{b} := \{\mathbf{b}_1, \dots, \mathbf{b}_{\bar{T}}\}$. Für $T \sim \text{PE}(\lambda, \mathbf{b})$ gilt insbesondere:

$$\text{pr}(T \in [\mathbf{b}_{t-1}, \mathbf{b}_t)) = S(\mathbf{b}_t) \cdot [1 - \exp(\mathbf{w}_t \cdot \lambda_t)] \quad (3.19)$$

$S(t)$ ist dabei die Survivor-Funktion aus (2.1). Der Beweis von (3.19) findet sich in Anhang B. Sei \bar{U} wie bisher die Anzahl der Untersuchungseinheiten und T_U die Zufallsvariable, die die Verweildauer der Untersuchungseinheit U angibt. Dann ist das piecewise exponential Modell gegeben durch das

- Beobachtungsmodell

$$T_U \sim \text{PE}(\lambda_U, b), \lambda_U = (\lambda_{1,U}(\alpha_1), \dots, \lambda_{\bar{T},U}(\alpha_{\bar{T}}), \lambda_{\bar{T}+1,U})',$$

$$\lambda_{tU}(\alpha_t) = \exp(\mathbf{z}'_U \cdot \alpha_t), \quad U = 1, \dots, \bar{U}, \quad t = 1, \dots, \bar{T}. \quad (3.20)$$

$\mathbf{z}_U \in \mathbb{R}^n$ ist der Design-Vektor der Untersuchungseinheit U , der die zeitkonstanten Kovariablen \mathbf{x}_U enthalten kann. Insbesondere nimmt man hier also an, daß die Beobachtung T_U piecewise exponential-verteilt ist. Gamerman (1991) leitet für dieses Modell die folgende Likelihood her:

$$L(\alpha_t) := \prod_{t=1}^{\bar{T}} \prod_{U \in R_t} L_{tU}(\alpha_t), \quad L_{tU}(\alpha_t) = (\lambda_{tU}(\alpha_t))^{y_{tU}} \cdot \exp(-\lambda_{tU}(\alpha_t) \cdot \Delta_{tU}), \quad (3.21)$$

wobei R_t die Risikomenge für das t -te Zeitintervall $[b_{t-1}, b_t)$ ist, t_U der beobachtete Wert von T_U und t_U ohne Einschränkung $\in [b_{t-1}, b_t)$, das heißt t_U ist die beobachtete „Überlebenszeit“ der Untersuchungseinheit U im t -ten Zeitintervall, $\Delta_{tU} := \max\{0, \min(w_t, t_U - b_{t-1})\}$ und $y_{tU} \in \{0, 1\}$ wie in Abschnitt 2.5. Wir nehmen an, daß Δ_{tU} nicht von α_t abhängt. Dann gilt für den Likelihood-Beitrag von $U \in R_t$:

$$\begin{aligned} L_{tU}(\alpha_t) &= \frac{(\lambda_{tU}(\alpha_t) \cdot \Delta_{tU})^{y_{tU}}}{\Delta_{tU}^{y_{tU}}} \cdot \exp(-\lambda_{tU}(\alpha_t) \cdot \Delta_{tU}) \propto \\ &\propto \frac{(\lambda_{tU}(\alpha_t) \cdot \Delta_{tU})^{y_{tU}}}{y_{tU}!} \cdot \exp(-\lambda_{tU}(\alpha_t) \cdot \Delta_{tU}). \end{aligned} \quad (3.22)$$

Da $(\bar{J}=1) \Rightarrow (y_{tU} \in \{0, 1\}) \Rightarrow (y_{tU}! = 1)$. Auf der rechten Seite von (3.22) steht die Poisson-Likelihood mit Parameter $\lambda_{tU}(\alpha_t) \cdot \Delta_{tU}$ und Beobachtung y_{tU} . Wählt man also die Poisson-Verteilung im

- Beobachtungsmodell

$$y_{tU} | \mathbf{r}_t^*, y_{t-1}^*, \mathbf{x}_U, \alpha_t \sim p(y_{tU} | \mathbf{r}_t^*, y_{t-1}^*, \mathbf{x}_U, \alpha_t) = P(\lambda_{tU}(\alpha_t) \cdot \Delta_{tU})$$

$$\lambda_{tU}(\alpha_t) = \exp(\eta_{tU}), \quad \eta_{tU} = \mathbf{z}'_U \cdot \alpha_t, \quad U = 1, \dots, \bar{U}, \quad t = 1, \dots, \bar{T}, \quad (3.23)$$

so erhält man eine Likelihood, die bezüglich α_t proportional ist zu $L_{tU}(\alpha_t)$. Für den Parameterübergang von α_{t-1} nach α_t behalten wir das lineare Übergangsmodell mit Normalverteilung (3.7) bei. Weil die Poisson-Verteilung in (3.23) der Klasse der einfachen Exponentialfamilie mit Gewichtsparameter $\omega_{tU} = 1/\Delta_{tU}$ angehört, ist (3.23) zusammen mit (3.7) ein verallgemeinertes Zustandsraummodell mit Responsefunktion $\mathbf{h}(\eta_{tU}) = \exp(\eta_{tU})$ und linearem Prädiktor $\eta_{tU} = \mathbf{z}'_U \cdot \alpha_t$. Wegen (3.22) ist (3.23), (3.7) auch die Zustandsraumform des piecewise exponential Modells (3.20).

Gamerman (1991) schlägt vor, im Übergangsmodell (3.7) anstatt der Kovarianzmatrix Q_t für die Fehlerterme ξ_t die Matrix $\mathbf{w}_t \cdot Q_t$ zu verwenden. Das würde bedeuten, daß im Falle einer großen Zeitintervalllänge auch eine große Varianz und damit starke Veränderungen der Regressionskoeffizienten beim Übergang von α_{t-1} nach α_t zugelassen werden. In Abschnitt 4.1 werden wir sehen, daß auch die Wahl von $(1/\mathbf{w}_t) \cdot Q_t$ sinnvoll sein kann. Mit beiden Ansätzen läßt sich der Fall behandeln, bei dem die Daten mit unterschiedlichen Zeitintervalllängen erhoben oder modelliert werden. Bei der Analyse der Magenkrebsdaten aus Teilabschnitt 1.1.1 haben wir unterschiedliche Zeitintervalllängen berücksichtigt und die Ergebnisse untereinander verglichen, die man mit der Wahl verschiedener, von Gamerman (1991) für diesen Datensatz vorgeschlagener Zeitgrids b erhält.

In diesem Abschnitt waren bisher die Kovariablen x_U zeitkonstant. Die Likelihood in (3.21) kann ebenso für zeitvariierende, externe Kovariable x_{tU} hergeleitet werden. Für zeitabhängige, interne Kovariable ist der Zusammenhang zwischen der Survivor-Funktion $S(t|\mathbf{x}_{tU}^*)$ und der Hazardrate $\lambda(t|\mathbf{x}_{tU}^*)$ nicht mehr in der einfachen Form wie in (2.3) gegeben. Fahrmeir und Tutz (1994, Teilabschnitt 9.2.4) zeigen, daß in diesem Fall $L_{tU}(\alpha_t)$ aus (3.21) beziehungsweise (3.22) als Partial-Likelihood-Beitrag und $L(\alpha_t)$ aus (3.21) als Partial-Likelihood verwendet werden kann.

Eine Verallgemeinerung des piecewise exponential Modell und die sich daraus ergebende Likelihood auf den Mehr-Zustands-Fall ist in Gamerman (1992, Abschnitt 6.2) enthalten.

4 Posteriori-Modus-Schätzung in dynamischen Modellen

In dieser Arbeit wollen wir das Glättungsproblem im Rahmen von Zustandsraummodellen besprechen. Dabei gehen wir zunächst auf den Zeitreihenfall ein, bevor wir in Abschnitt 4.1 den Ein-Episoden- und in Abschnitt 4.2 den Mehr-Episoden-Fall der Ereignisanalyse behandeln. Das Glättungsproblem besteht darin, bei gegebenen Beobachtungen $y_1, \dots, y_{\bar{T}}$ und Hyperparametern $a_0, Q_0, Q_1, Q_2, \dots, Q_{\bar{T}}$, die nicht beobachtbaren, latenten Zustände α_t , $t = 1, \dots, \bar{T}$, zu schätzen. Für das lineare Zustandsraummodell (3.1), (3.2) ist die bezüglich des „mean square error“ optimale Lösung der Posteriori-Erwartungswert

$$a_{t|\bar{T}} = E(\alpha_t | y_{\bar{T}}^*, x_{\bar{T}}^*), \quad t = 1, \dots, \bar{T}, \quad (4.1)$$

(siehe dazu Anderson und Moore, 1979, Theorem 3.1 in Abschnitt 2.3), wobei hier wegen der Normalverteilungsannahme Erwartungswert und Modus übereinstimmen. Im Falle des verallgemeinerten Zustandsraummodells (3.6), (3.7) basiert die Schätzung auf der Posteriori-Dichte $p(\alpha_{\bar{T}}^* | y_{\bar{T}}^*, x_{\bar{T}}^*)$. Die Posteriori-Verteilung ist beim linearen Zustandsraummodell wieder die Normalverteilung, so daß die Posteriori-Dichte analytisch zugänglich ist. Im allgemeinen ist dies jedoch nicht der Fall.

Man kann zwischen drei Schätzansätzen für das verallgemeinerte Zustandsraummodell unterscheiden. Eine Möglichkeit besteht in der Verwendung einer konjugierten Priori-Verteilung, um die notwendigen Integrationen analytisch durchführen zu können (vergleiche dazu West, Harrison und Migon, 1985, und West und Harrison, 1989). Die Anwendung des Gibbs Samplers wie in Carlin, Polson und Stoffer (1992) und von Monte-Carlo-Techniken ist eine weitere Alternative. Hier eröffnet sich der Weg zur Berechnung der posteriori-Momente Erwartungswert und Varianz oder sogar zu einer kompletten Bayes-Analyse. Einen Überblick findet man beispielsweise in Fahrmeir und Tutz (1994, Teilabschnitt 8.3.2 und Appendix A.5), in Kass und Raftery (1994, §§ 4 und 6), und für Markov-Ketten-Monte-Carlo-Methoden in Tierney (1994). Die Anwendung dieser Verfahren im Bereich der verallgemeinerten Zustandsraummodelle und methodische Erweiterungen werden in Knorr-Held (1995) besprochen. Der dritte Ansatz, bei dem das Problem der Integration vollkommen umgangen wird, ist das Posteriori-Modus-Schätzkonzept. Darauf wollen wir im folgenden näher eingehen und die Verfahren und Algorithmen entwickeln, die zur Schätzung und Analyse der Datensätze aus Abschnitt 1.1 verwendet wurden.

Der Posteriori-Modus-Schätzer $\mathbf{a} \equiv (\mathbf{a}'_{0|\bar{T}}, \mathbf{a}'_{1|\bar{T}}, \dots, \mathbf{a}'_{\bar{T}|\bar{T}})' \in \mathbb{R}^m$ mit $m := (\bar{T} + 1)n$ ist definiert als

$$\mathbf{a} \equiv (\mathbf{a}'_{0|\bar{T}}, \mathbf{a}'_{1|\bar{T}}, \dots, \mathbf{a}'_{\bar{T}|\bar{T}})' := \arg \max_{\alpha_{\bar{T}}^*} \left\{ p(\alpha_{\bar{T}}^* | \mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*) \right\}, \quad (4.2)$$

das heißt \mathbf{a} ist der gemeinsame Modus der Posteriori-Verteilung. Das Ziel ist es nun, die Funktion $p(\alpha_{\bar{T}}^* | \mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*)$ in Abhängigkeit von $\alpha_{\bar{T}}^*$ zu maximieren. Die wiederholte Anwendung der Definition der bedingten Wahrscheinlichkeit liefert

$$\begin{aligned} p(\alpha_{\bar{T}}^* | \mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*) &= \\ &= \frac{1}{p(\mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*)} \left\{ \prod_{t=1}^{\bar{T}} p(\mathbf{y}_t | \alpha_t^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*) \prod_{t=1}^{\bar{T}} p(\mathbf{x}_t | \alpha_{t-1}^*, \mathbf{y}_{t-1}^*, \mathbf{x}_{t-1}^*) \prod_{t=1}^{\bar{T}} p(\alpha_t | \alpha_{t-1}^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*) \cdot p(\alpha_0) \right\}. \end{aligned} \quad (4.3)$$

Mit den Voraussetzungen (3.8), (3.9), (3.10) und weil $p(\mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*)$ nicht von $\alpha_{\bar{T}}^*$ abhängt, gilt

$$p(\alpha_{\bar{T}}^* | \mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*) \propto \prod_{t=1}^{\bar{T}} p(\mathbf{y}_t | \alpha_t^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*) \prod_{t=1}^{\bar{T}} p(\alpha_t | \alpha_{t-1}^*) \cdot p(\alpha_0).$$

Indem wir auf der rechten Seite logarithmieren und die Normalverteilungsdichten aus dem Übergangsmodell (3.7) einsetzen, erhalten wir eine Funktion $PL: \mathbb{R}^m \rightarrow \mathbb{R}$, die wir als penalisierte Loglikelihood bezeichnen wollen:

$$\begin{aligned} PL(\alpha_{\bar{T}}^*) &:= \sum_{t=1}^{\bar{T}} \ln p(\mathbf{y}_t | \alpha_t^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*) - \frac{1}{2} (\alpha_0 - \mathbf{a}_0)' \mathbf{Q}_0^{-1} (\alpha_0 - \mathbf{a}_0) \\ &\quad - \frac{1}{2} \sum_{t=1}^{\bar{T}} (\alpha_t - \mathbf{F}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} (\alpha_t - \mathbf{F}_t \alpha_{t-1}). \end{aligned} \quad (4.4)$$

Die Dichten $p(\mathbf{y}_t | \alpha_t^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*)$ sind dabei durch die Exponentialfamilienverteilungen im Beobachtungsmodell (3.6) definiert.

Damit gilt:

$$\mathbf{a} \equiv (\mathbf{a}'_{0|\bar{T}}, \mathbf{a}'_{1|\bar{T}}, \dots, \mathbf{a}'_{\bar{T}|\bar{T}})' = \arg \max_{\alpha_{\bar{T}}^*} \left\{ PL(\alpha_{\bar{T}}^*) \right\}. \quad (4.5)$$

Die Berechnung des Posteriori-Modus kann daher durch die Maximierung der penalisierten Loglikelihood (4.4) bezüglich $\alpha_{\bar{T}}^*$ erfolgen.

Die nichtlineare Funktion $PL(\alpha_{\bar{T}}^*)$ in (4.4) setzt sich aus zwei Teilen zusammen. Aus Bayes'scher Sicht ist der erste Term die Loglikelihood aus dem Beobachtungsmodell (3.6), und die letzten beiden Terme sind Glatheitspriori-Dichten, die durch das Übergangsmodell (3.7) definiert sind. Vom nonparametrischen Standpunkt aus sind die Zustände $\alpha_{\bar{T}}^*$ feste, aber unbekannte und zu schätzende Parameter. Der erste Term stellt dann ein Abstandsmaß zwischen den Daten und der Schätzung dar. Die letzten beiden Teile in (4.4) sind Strafterme, die zu starke Schwankungen und Veränderungen beim Parameterübergang von α_{t-1} nach α_t , $t = 1, \dots, \bar{T}$, bestrafen. Das Gewicht, mit dem bestraft wird, ist durch Q_0^{-1} und Q_t^{-1} , $t = 1, \dots, \bar{T}$, gegeben. Dies läßt sich an einem einfachen Trendmodell ($\dim(\alpha_t) = 1$) mit RW1 für den Parameterübergang und $Q_0 = q_0$, $Q_1 = Q_2 = \dots = Q_{\bar{T}} = q$ veranschaulichen. Die penalisierte Loglikelihood ist dann gegeben durch

$$PL(\alpha_{\bar{T}}^*) := \sum_{t=1}^{\bar{T}} \ln p(y_t | \alpha_t, y_{t-1}^*, x_t^*) - \frac{1}{2 \cdot q_0} (\alpha_0 - a_0)^2 - \frac{1}{2 \cdot q} \sum_{t=1}^{\bar{T}} (\alpha_t - \alpha_{t-1})^2, \quad (4.6)$$

und die Werte $1/q_0$ und $1/q$ übernehmen die Rolle von Glättungsparametern.

Für den Spezialfall des linearen Zustandsraummodells (3.1), (3.2) lautet die penalisierte Loglikelihood

$$PL(\alpha_{\bar{T}}^*) := -\frac{1}{2} \sum_{t=1}^{\bar{T}} (y_t - Z_t \alpha_t)' \Sigma_t^{-1} (y_t - Z_t \alpha_t) - \frac{1}{2} (\alpha_0 - a_0)' Q_0^{-1} (\alpha_0 - a_0) - \frac{1}{2} \sum_{t=1}^{\bar{T}} (\alpha_t - F_t \alpha_{t-1})' Q_t^{-1} (\alpha_t - F_t \alpha_{t-1}). \quad (4.7)$$

(4.7) ist ein Kleinst-Quadrat-Kriterium mit Strafterm. Das nichtlineare Optimierungsproblem in (4.5) reduziert sich in diesem Falle auf ein quadratisches Optimierungsproblem. Weil der posteriori-Modus und der posteriori-Erwartungswert bei der Normalverteilung übereinstimmen, kann für lineare Zustandsraummodelle zur effizienten Lösung des quadratischen Optimierungsproblems (4.5) der lineare Kalman Filter und Glätter (KFG) verwendet werden. Das Verfahren wird in Abschnitt 5.1 besprochen.

Zur Lösung des nichtlinearen Optimierungsproblems (4.5) für das verallgemeinerte Zustandsraummodell (3.6), (3.7) kann prinzipiell jeder Algorithmus aus der nichtlinearen

Optimierung verwendet werden. Eine Einführung in dieses Gebiet der numerischen Mathematik findet man in Werner (1992) und einen Überblick beispielsweise in Gill, Murray und Wright (1981). Wir wollen hier das Fisher-Scoring-Verfahren verwenden, das häufig bei verallgemeinerten linearen Modellen eingesetzt wird. Dazu zeigen wir in Abschnitt 5.1, daß auch die Fisher-Scoring-Iterationen auf effiziente Weise mit einem modifizierten KFG-Algorithmus ausgeführt werden können.

Bevor wir das lineare Gleichungssystem für eine Fisher-Scoring-Iteration herleiten, schreiben wir die penalisierte Loglikelihood (4.4) in Matrixnotation:

$$PL(\alpha_{\bar{T}}^*) = l(\alpha_{\bar{T}}^*) - \frac{1}{2}(\alpha_{\bar{T}}^*)' P(\alpha_{\bar{T}}^*), \quad (4.8)$$

mit

$$l(\alpha_{\bar{T}}^*) = \sum_{t=0}^{\bar{T}} l_t(\alpha_t), \quad l_t(\alpha_t) := \ln p(y_t | \alpha_t, y_{t-1}^*, x_t^*), \quad t = 1, \dots, \bar{T},$$

$$l_0(\alpha_0) := -(\alpha_0 - a_0)' Q_0^{-1} (\alpha_0 - a_0) / 2.$$

Die Penalty-Matrix $P \in \mathbb{R}^{m,m}$ ist symmetrisch und blocktridiagonal. Die Blöcke ergeben sich sofort aus (4.4) zu

$$P = \begin{bmatrix} P_{00} & P_{01} & & & 0 \\ P_{10} & P_{11} & P_{12} & & \\ & P_{21} & \ddots & \ddots & \\ & & \ddots & \ddots & P_{\bar{T}-1, \bar{T}} \\ 0 & & & P_{\bar{T}, \bar{T}-1} & P_{\bar{T}, \bar{T}} \end{bmatrix} \quad (4.9)$$

mit

$$P_{t-1,t} = -F_t' Q_t^{-1} \quad (t=1, \dots, \bar{T}),$$

$$P_{t,t-1} = P_{t-1,t}' \quad (t=1, \dots, \bar{T}),$$

$$P_{00} = F_1' Q_1^{-1} F_1,$$

$$P_{tt} = Q_t^{-1} + F_{t+1}' Q_{t+1}^{-1} F_{t+1} \quad (t=1, \dots, \bar{T}),$$

$$F_{T+1} = 0_{n,n}.$$

Um auch einen Fisher-Scoring-Schritt in kompakter Matrixnotation darzustellen, definieren wir die folgenden Vektoren und Matrizen: In

$$y' := (a_0', y_1', \dots, y_{\bar{T}}') \quad (4.10)$$

fassen wir sämtliche Beobachtungen und $\mathbf{a}_0 = \mathbf{E}(\alpha_0)$ zusammen, in

$$\mu(\alpha_{\bar{T}}^*)' := \{\alpha_0', \mu_1'(\alpha_1), \dots, \mu_{\bar{T}}'(\alpha_{\bar{T}})\} \quad (4.11)$$

den zu schätzenden Parametervektor α_0 und die Erwartungswerte $\mu_t(\alpha_t) = h(Z_t \alpha_t)$ aus dem Beobachtungsmodell (3.6) in Abhängigkeit von den zu schätzenden Zustandsvektoren α_t , $t = 1, \dots, \bar{T}$, in der blockdiagonalen Matrix

$$\Sigma(\alpha_{\bar{T}}^*) := \text{diag}\{\mathbf{Q}_0, \Sigma_1(\alpha_1), \dots, \Sigma_{\bar{T}}(\alpha_{\bar{T}})\} \quad (4.12)$$

die Kovarianzmatrizen aus dem Beobachtungsmodell (3.6) und der Start-Priori-Verteilung, in der Rechtecksmatrix

$$\mathbf{Z} := \begin{pmatrix} \mathbf{I}_n & & & \mathbf{0} \\ & \mathbf{Z}_1 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{Z}_{\bar{T}} \end{pmatrix} \quad (4.13)$$

die Design-Matrizen und in der blockdiagonalen Matrix

$$\mathbf{D}(\alpha_{\bar{T}}^*) = \text{diag}\{\mathbf{I}_n, \mathbf{D}_1(\alpha_1), \dots, \mathbf{D}_{\bar{T}}(\alpha_{\bar{T}})\} \quad (4.14)$$

die Jacobi-Matrizen $\mathbf{D}_t(\alpha_t) := \partial h(\eta_t)/\partial \eta$ der Response-Funktion $h(\eta)$ ausgewertet am linearen Prädiktor $\eta_t = Z_t \alpha_t$.

Weil wir im Beobachtungsmodell (3.6) eine Verteilung aus der Klasse der einfachen Exponentialfamilie angenommen haben, gilt für die Score-Funktion von $l(\alpha_t)$:

$$\begin{aligned} s_t(\alpha_t) &= \mathbf{Z}_t' \mathbf{D}_t(\alpha_t) \Sigma_t^{-1}(\alpha_t) \{y_t - \mu_t(\alpha_t)\}, \quad t = 1, \dots, \bar{T}, \\ s_0(\alpha_0) &= \mathbf{Q}_0^{-1}(\mathbf{a}_0 - \alpha_0). \end{aligned} \quad (4.15)$$

Eine Herleitung findet man beispielsweise in Fahrmeir und Tutz (1994, Appendix A.1).

Damit erhalten wir die Score-Funktion von $l(\alpha_{\bar{T}}^*)$ in (4.8):

$$s(\alpha_{\bar{T}}^*) = \{s_0'(\alpha_0), s_1'(\alpha_1), \dots, s_{\bar{T}}'(\alpha_{\bar{T}})\}' := \mathbf{Z}' \mathbf{D}(\alpha_{\bar{T}}^*) \Sigma^{-1}(\alpha_{\bar{T}}^*) \{y - \mu(\alpha_{\bar{T}}^*)\}. \quad (4.16)$$

Nach (4.15) ergibt sich dann für die erwarteten Informationsmatrizen von $l(\alpha_t)$:

$$\mathbf{S}_t(\alpha_t) = \mathbf{Z}_t' \mathbf{W}_t(\alpha_t) \mathbf{Z}_t, \quad t = 1, \dots, \bar{T},$$

$$S_0 = Q_0^{-1}, \quad (4.17)$$

mit den Gewichtsmatrizen

$$\begin{aligned} W_t(\alpha_t) &= D_t(\alpha_t) \Sigma_t^{-1}(\alpha_t) D_t'(\alpha_t), \quad t = 1, \dots, \bar{T}, \\ W_0 &= Q_0^{-1}. \end{aligned} \quad (4.18)$$

Fassen wir diese Matrizen in der „großen“ blockdiagonalen Gewichtsmatrix

$$W(\alpha_{\bar{T}}^*) = \text{diag}\{W_0, W_1(\alpha_1), \dots, W_{\bar{T}}(\alpha_{\bar{T}})\} := D(\alpha_{\bar{T}}^*) \Sigma^{-1}(\alpha_{\bar{T}}^*) D'(\alpha_{\bar{T}}^*) \quad (4.19)$$

zusammen, so können wir die erwartete Informationsmatrix $S(\alpha_{\bar{T}}^*)$ von $l(\alpha_{\bar{T}}^*)$ in (4.8) ebenfalls in Form einer Blockdiagonalmatrix schreiben:

$$S(\alpha_{\bar{T}}^*) = \text{diag}\{S_0, S_1(\alpha_1), \dots, S_{\bar{T}}(\alpha_{\bar{T}})\} := Z' W(\alpha_{\bar{T}}^*) Z. \quad (4.20)$$

Die erste Ableitung der penalisierten Loglikelihood $PL(\alpha_{\bar{T}}^*)$ in (4.4) nach $\alpha_{\bar{T}}^*$ ist

$$g(\alpha_{\bar{T}}^*) := \partial PL(\alpha_{\bar{T}}^*) / \partial \alpha_{\bar{T}}^* = s(\alpha_{\bar{T}}^*) - P \cdot \alpha_{\bar{T}}^*, \quad (4.21)$$

und die blocktridiagonale erwartete Informationsmatrix von $PL(\alpha_{\bar{T}}^*)$ ist durch

$$G(\alpha_{\bar{T}}^*) := -E\{\partial^2 PL(\alpha_{\bar{T}}^*) / \partial(\alpha_{\bar{T}}^*) \partial(\alpha_{\bar{T}}^*)'\} = S(\alpha_{\bar{T}}^*) + P = Z' W(\alpha_{\bar{T}}^*) Z + P \quad (4.22)$$

gegeben.

Der Fisher-Scoring-Algorithmus ist eine allgemeine Gradientenmethode, deren Folge von Iterationspunkten $\{\mathbf{a}^{(i)}\}_{i=1}^{\infty}$ unter gewissen Voraussetzungen (siehe dazu beispielsweise Werner, 1992) gegen die Lösung, den Posteriori-Modus-Schätzer $\mathbf{a} \in \mathbb{R}^m$ in (4.5), konvergiert. Für unser Optimierungsproblem in (4.5) ist ein Fisher-Scoring-Schritt von einem Iterationspunkt $\mathbf{a}^{(i)} = ((\mathbf{a}_0^{(i)})', (\mathbf{a}_1^{(i)})', \dots, (\mathbf{a}_{\bar{T}}^{(i)})')' \in \mathbb{R}^m$ nach $\mathbf{a}^{(i+1)} \in \mathbb{R}^m$ gegeben durch

$$G(\mathbf{a}^{(i)}) \{\mathbf{a}^{(i+1)} - \mathbf{a}^{(i)}\} = g(\mathbf{a}^{(i)}). \quad (4.23)$$

Dies läßt sich unter Verwendung von (4.16), (4.19), (4.21) und (4.22) algebraisch umformen zu

$$\mathbf{a}^{(i+1)} = \{Z' W(\mathbf{a}^{(i)}) Z + P\}^{-1} Z' W(\mathbf{a}^{(i)}) \tilde{\mathbf{y}}(\mathbf{a}^{(i)}) \quad (4.24)$$

mit der „working observation“

$$\tilde{\mathbf{y}}(\mathbf{a}^{(i)}) = \left\{ \tilde{\mathbf{y}}'_0, \tilde{\mathbf{y}}'_1(\mathbf{a}_1^{(i)}), \dots, \tilde{\mathbf{y}}'_T(\mathbf{a}_T^{(i)}) \right\}' := \left\{ \mathbf{D}^{-1}(\mathbf{a}^{(i)}) \right\}' \left\{ \mathbf{y} - \boldsymbol{\mu}(\mathbf{a}^{(i)}) \right\} + \mathbf{Z}\mathbf{a}^{(i)}. \quad (4.25)$$

Die Komponenten dieses Vektors sind dabei definiert durch $\tilde{\mathbf{y}}_0 := \mathbf{a}_0$ und $\tilde{\mathbf{y}}_t(\mathbf{a}_t^{(i)}) :=$

$$\left\{ \mathbf{D}_t^{-1}(\mathbf{a}_t^{(i)}) \right\}' \left\{ \mathbf{y}_t - \boldsymbol{\mu}_t(\mathbf{a}_t^{(i)}) \right\} + \mathbf{Z}_t \mathbf{a}_t^{(i)}, \quad t = 1, \dots, T.$$

Für das lineare Zustandsraummodell (3.1), (3.2) mit der penalisierten Loglikelihood $PL(\boldsymbol{\alpha}_{\bar{T}}^*)$ aus (4.7) reduziert sich (4.24) zu

$$\mathbf{a} = (\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{Z} + \mathbf{P})^{-1}\mathbf{Z}'\boldsymbol{\Sigma}^{-1}\mathbf{y}, \quad (4.26)$$

denn hier ist $\boldsymbol{\mu}(\boldsymbol{\alpha}_{\bar{T}}^*) = \mathbf{Z}\boldsymbol{\alpha}_{\bar{T}}^*$ und $\mathbf{D}(\boldsymbol{\alpha}_{\bar{T}}^*)$ die identische Matrix. Ferner ist die Score-Funktion von $l(\boldsymbol{\alpha}_{\bar{T}}^*)$ mit $\boldsymbol{\Sigma} := \text{diag}(\mathbf{Q}_0, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_{\bar{T}})$, $\boldsymbol{\Sigma}_t = \text{var}(\mathbf{y}_t | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \boldsymbol{\alpha}_t)$, $t = 1, \dots, \bar{T}$ gegeben durch

$$\mathbf{s}(\boldsymbol{\alpha}_{\bar{T}}^*) = \mathbf{Z}'\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{Z}\boldsymbol{\alpha}_{\bar{T}}^*).$$

Die Gewichtsmatrix $\mathbf{W}(\boldsymbol{\alpha}_{\bar{T}}^*)$ wird zu $\boldsymbol{\Sigma}^{-1}$, und die „working observation“ $\tilde{\mathbf{y}}(\boldsymbol{\alpha}_{\bar{T}}^*)$ aus (4.25) zur tatsächlichen Beobachtung \mathbf{y} . Zur Lösung des linearen Gleichungssystems in (4.26) kann der KFG-Algorithmus eingesetzt werden. Die Lösungsvektoren, die man mit diesem Verfahren nach dem Glättungsschritt erhält, sind die Schätzungen $\mathbf{a}_{t|\bar{T}}$, $t = 0, \dots, \bar{T}$, mit

$$\mathbf{a} = (\mathbf{a}'_{0|\bar{T}}, \mathbf{a}'_{1|\bar{T}}, \dots, \mathbf{a}'_{\bar{T}|\bar{T}})'. \quad (4.27)$$

Was bei diesem Posteriori-Modus-Schätzansatz im Hinblick auf die Ereignisanalyse zu beachten ist, zeigen die nächsten beiden Abschnitte.

4.1 Ein-Episoden-Fall

In diesem Abschnitt betrachten wir den Ein-Episoden-Fall $\bar{K} = 1$, zu dem wir die Ein-Episoden-Ein-Zustands- und die Mehr-Zustands-Situation zusammenfassen. Die Kovariablen seien zunächst entweder deterministisch oder extern im Sinne von Kalbfleisch und Prentice (1980).

Es sei J_U die Zufallsvariable die angibt, in welchen Zustand die Untersuchungseinheit U im Zeitintervall t_U wechselt. t_U ist dabei der Wert der Zufallsvariablen T_U , die wie in Abschnitt 3.2 die Verweildauer von U angibt. Für die gewöhnliche Likelihood L_G , die unter den in Abschnitt 2.4 besprochenen Voraussetzungen „random censoring“ und „nichtinformativ“ Zensierung durch

$$L_G = \prod_{U=1}^{\bar{U}} \text{pr}(T_U = t_U, J_U = j_U | \mathbf{x}_t^*)^{\delta_U} \cdot S(t_U | \mathbf{x}_t^*)^{1-\delta_U} \quad (4.28)$$

mit δ_U aus (2.22) und $S(t)$ aus (2.1) gegeben ist, leiten Fahrmeir und Tutz (1994, Teilabschnitt 9.3.2) den folgenden Zusammenhang her:

$$L_G = \prod_{t=1}^{\bar{T}} \prod_{U \in R_t} L_{tU}(\alpha_t) \quad \text{mit} \quad L_{tU}(\alpha_t) := \left\{ \left(1 - \sum_{j=1}^{\bar{J}} \lambda_j(t | \mathbf{x}_t^*) \right)^{1 - \sum_{j=1}^{\bar{J}} y_{tUj}} \cdot \prod_{j=1}^{\bar{J}} \lambda_j(t | \mathbf{x}_t^*)^{y_{tUj}} \right\}. \quad (4.29)$$

Die Abhängigkeit von α_t in (4.29) ist durch den Regressionsansatz (3.12) beziehungsweise (3.23) gegeben. Der Likelihoodbeitrag $L_{tU}(\alpha_t)$ in (4.29) hat die Gestalt einer Multinomialverteilungslikelihood. Den gleichen Beitrag erhalten wir für unser multinomiales Zustandsraummodell aus Abschnitt 3.1, wenn wir annehmen, daß

$$p(y_{tU} | y_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*; \alpha_t) = p(y_{tU} | y_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*), \quad \forall U \notin R_t, \quad t = 1, \dots, \bar{T}, \quad (4.30)$$

mit den Größen y_t^* , \mathbf{r}_t^* wie in (2.28) und \mathbf{x}_t^* wie in (2.31). (4.30) bedeutet, daß α_t gegeben $y_{t-1}^*, \mathbf{x}_t^*$ und \mathbf{r}_t^* keinen Einfluß auf y_{tU} und damit auf einen Zustandswechsel hat, wenn sich U nicht mehr unter Risiko befindet. Diese Annahme ist sinnvoll, denn für $U \notin R_t$

gilt im Ein-Episoden-Fall: Es hat bereits ein Zustandswechsel stattgefunden und ein weiterer kann nicht eintreten, oder es ist eine Zensierung aufgetreten und ein möglicher Zustandswechsel kann nicht beobachtet werden. Ferner läßt sich (4.30) mit (4.29) rechtfertigen.

Außerdem nehmen wir an, daß die Beobachtungen y_{tU} gegeben $y_{t-1}^*, x_t^*, r_t^*, \alpha_t$ für $U = 1, \dots, \bar{U}$ bedingt unabhängig sind:

$$p(y_t | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) = \prod_{U=1}^{\bar{U}} p(y_{tU} | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) \quad , \quad t = 1, \dots, \bar{T} \quad , \quad (4.31)$$

mit y_t wie in (2.28) angegeben. (4.31) entspricht der bei reinen Querschnittsdaten gängigen Unabhängigkeitsannahme. (4.30) und (4.31) zusammen ergeben

$$\begin{aligned} p(y_t | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) &\stackrel{(4.31)}{=} \prod_{U=1}^{\bar{U}} p(y_{tU} | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) = \\ &= \prod_{U \in R_t} p(y_{tU} | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) \cdot \prod_{U \notin R_t} p(y_{tU} | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) \stackrel{(4.30)}{\propto} \\ &\propto \prod_{U \in R_t} p(y_{tU} | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) \quad , \quad t = 1, \dots, \bar{T} \quad . \end{aligned} \quad (4.32)$$

Die Proportionalität in (4.32) und auch im folgenden ist stets bezüglich den Zuständen α_t aus dem Regressionsansatz von (3.12) oder (3.23) zu verstehen.

Ferner benötigen wir noch die folgenden Annahmen, die weitgehend bereits im Zeitreihenfall gemacht wurden. Dabei sei $\alpha_t^* := (\alpha'_0, \alpha'_1, \dots, \alpha'_t)'$ wie in (3.8):

$$\begin{aligned} p(x_t, r_t | y_{t-1}^*, x_{t-1}^*, r_{t-1}^*; \alpha_{t-1}^*) &= p(x_t, r_t | y_{t-1}^*, x_{t-1}^*, r_{t-1}^*) \quad , \\ p(y_t | y_{t-1}^*, x_t^*, r_t^*; \alpha_t^*) &= p(y_t | y_{t-1}^*, x_t^*, r_t^*; \alpha_t) \quad \text{und} \\ p(\alpha_t | y_{t-1}^*, x_t^*, r_t^*; \alpha_{t-1}^*) &= p(\alpha_t | \alpha_{t-1}) \quad , \quad t = 1, \dots, \bar{T} \quad . \end{aligned} \quad (4.33)$$

Die erste Annahme in (4.33) gilt bei „nichtinformativer Zensierung“ und bei einem externen Kovariablenprozeß $\{x_t\}$, $t = 1, \dots, \bar{T}$. Kalbfleisch und Prentice (1980) unterscheiden drei Arten von externen Kovariablen, zu denen auch die fixen, zeitinvarianten Kovariablen zählen.

Mit den Annahmen (4.32) und (4.33) läßt sich der Posteriori-Modus

$$\mathbf{a} = (\mathbf{a}'_{0,\bar{T}}, \mathbf{a}'_{1,\bar{T}}, \dots, \mathbf{a}'_{\bar{T},\bar{T}})' := \arg \max_{\alpha_{\bar{T}}^*} \{p(\alpha_{\bar{T}}^* | \mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*, \mathbf{r}_{\bar{T}}^*)\} \quad (4.34)$$

wieder dadurch berechnen, daß man eine penalisierte Likelihood $\text{PL}(\alpha_{\bar{T}}^*)$ mit

$$\begin{aligned} \text{PL}(\alpha_{\bar{T}}^*) &= \sum_{t=1}^{\bar{T}} \sum_{U \in R_t} \ln p(\mathbf{y}_{tU} | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*, \alpha_t) - \frac{1}{2} (\alpha_0 - \mathbf{a}_0)' \mathbf{Q}_0^{-1} (\alpha_0 - \mathbf{a}_0) \\ &\quad - \frac{1}{2} \sum_{t=1}^{\bar{T}} (\alpha_t - \mathbf{F}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} (\alpha_t - \mathbf{F}_t \alpha_{t-1}) \end{aligned} \quad (4.35)$$

bezüglich $\alpha_{\bar{T}}^*$ maximiert, denn

$$\begin{aligned} p(\alpha_{\bar{T}}^* | \mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*, \mathbf{r}_{\bar{T}}^*) &= \\ &= \frac{1}{p(\mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*, \mathbf{r}_{\bar{T}}^*)} \left\{ \prod_{t=1}^{\bar{T}} p(\mathbf{y}_t | \alpha_t^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*) \prod_{t=1}^{\bar{T}} p(\mathbf{x}_t, \mathbf{r}_t | \alpha_{t-1}^*, \mathbf{y}_{t-1}^*, \mathbf{x}_{t-1}^*, \mathbf{r}_{t-1}^*) \cdot \right. \\ &\quad \left. \prod_{t=1}^{\bar{T}} p(\alpha_t | \alpha_{t-1}^*, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*) \cdot p(\alpha_0) \right\} \stackrel{(4.32), (4.33)}{\propto} \\ &\propto \prod_{t=1}^{\bar{T}} \prod_{U \in R_t} p(\mathbf{y}_{tU} | \alpha_t, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*) \prod_{t=1}^{\bar{T}} p(\alpha_t | \alpha_{t-1}) \cdot p(\alpha_0). \end{aligned} \quad (4.36)$$

Logarithmieren und Einsetzen der Normalverteilungsdichte aus dem Übergangsmodell (3.7) ergibt, bis auf eine Normierungskonstante, (4.35). Im Falle unseres multinomialen Zustandsraummodells aus Abschnitt 3.1 ist für $p(\mathbf{y}_{tU} | \alpha_t, \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*)$ die Dichte der Multinomialverteilung aus (3.12) einzusetzen und wir erhalten

$$\begin{aligned} \text{PL}(\alpha_{\bar{T}}^*) &= \\ &= \sum_{t=1}^{\bar{T}} \sum_{U \in R_t} \ln \{L_{tU}(\alpha_t)\} - \frac{1}{2} (\alpha_0 - \mathbf{a}_0)' \mathbf{Q}_0^{-1} (\alpha_0 - \mathbf{a}_0) - \frac{1}{2} \sum_{t=1}^{\bar{T}} (\alpha_t - \mathbf{F}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} (\alpha_t - \mathbf{F}_t \alpha_{t-1}) \end{aligned} \quad (4.37)$$

mit $L_{tU}(\alpha_t)$ wie in (4.29). Für das piecewise exponential Modell (3.23), (3.7) mit der Wahl von $\text{cov}(\xi_t) = \mathbf{w}_t \cdot \mathbf{Q}_t$ bekommt man die folgende penalisierte Loglikelihood unter Berücksichtigung von (3.21):

$$\begin{aligned}
\text{PL}(\alpha_{\bar{T}}^*) = & \sum_{t=1}^{\bar{T}} \sum_{U \in R_t} \{y_{tU} \cdot \eta_{tU} - \Delta_{tU} \cdot \exp(\eta_{tU})\} - \frac{1}{2}(\alpha_0 - a_0)' Q_0^{-1}(\alpha_0 - a_0) \\
& - \frac{1}{2} \sum_{t=1}^{\bar{T}} \frac{1}{w_t} (\alpha_t - F_t \alpha_{t-1})' Q_t^{-1}(\alpha_t - F_t \alpha_{t-1}),
\end{aligned} \tag{4.38}$$

mit dem linearen Prädiktor $\eta_{tU} = z'_{tU} \cdot \alpha_t$. $t \in \{1, \dots, \bar{T}\}$ bezeichnet dabei jeweils die Nummer eines Zeitintervalls. Nach Definition von Δ_{tU} in (3.21) ist $\Delta_{tU} = w_t$ für $(U \in R_t) \wedge (y_{tU} = 0)$. Damit wird in (4.38) im Falle von großen Zeitintervalllängen das Beobachtungsmodell und damit die Daten stark gewichtet, verstärkt durch den Faktor $1/w_t$ in den Summen des Strafterms.

Eine andere Möglichkeit besteht in der Wahl von

$$\text{cov}(\xi_t) = \frac{1}{w_t} \cdot Q_t. \tag{4.39}$$

Die Bedeutung von (4.39) liegt darin, daß im Falle einer großen Zeitintervalllänge eine große Varianz beziehungsweise Schwankung beim Parameterübergang von α_{t-1} nach α_t mit starkem Gewicht bestraft wird. Da aus nonparametrischer Sicht, wie wir am Beispiel in (4.6) gesehen haben, der Kehrwert von $\text{cov}(\xi_t)$ als Maß für die Stärke der Glättung betrachtet werden kann, erfolgt bei (4.39) für große Zeitintervalllängen eine im Vergleich zu (4.38) stärkere Glättung der Schätzung. Dies kann im Rahmen der Ereignisanalyse sinnvoll eingesetzt werden, weil hier häufig in den ersten Zeitintervallen viele Ereignisse auftreten und damit eine große Zahl von Beobachtungen gemacht werden kann, während gegen Ende des Beobachtungszeitraumes die Datenlage sehr dünn und die Anzahl der Untersuchungseinheiten im Risikoset R_t sehr gering wird. Wählt man dann am Anfang des Beobachtungszeitraumes entsprechend der vielen Information, die in den Daten steckt, kurze Zeitintervalllängen und am Ende große Intervalllängen, um wenigstens eine kleine Zahl von Ereignissen aufzusammeln, so bewirkt (4.39) verglichen mit (4.38) zu Beginn, bei viel Dateninformation, ein niedrigeres Maß an Glättung und sorgt gegen Ende, bei geringer Datenlage, für die Existenz der Schätzer.

Für den Fall

$$w = w_1 = w_2 = \dots = w_{\bar{T}} \tag{4.40}$$

führt der Ansatz $\text{cov}(\xi_t) = \mathbf{w} \cdot \mathbf{Q}_t$ und der in (4.39) bei entsprechender Wahl von \mathbf{Q}_t zum gleichen Posteriori-Modus-Schätzergebnis.

Liegen Schätzungen für die Zustände α_t vor, so kann die Survivor-Funktion $\mathbf{S}(t_U | \mathbf{x}_U)$ im piecewise exponential Modell mit $t_U \in [\mathbf{b}_{t-1}, \mathbf{b}_t)$ wegen (2.3) und (3.18) nach folgender Formel bestimmt werden:

$$\mathbf{S}(t_U | \mathbf{x}_U) = \exp\left(-\Delta_{tU} \lambda_t - \sum_{s=1}^{t-1} \mathbf{w}_s \lambda_s\right) \stackrel{(3.23)}{=} \exp\left(-\Delta_{tU} \exp(\mathbf{z}_U \cdot \alpha_t) - \sum_{s=1}^{t-1} \mathbf{w}_s \exp(\mathbf{z}_U \cdot \alpha_s)\right). \quad (4.41)$$

Zur Lösung von (4.34) wollen wir wieder wie im Zeitreihenfall das Fisher-Scoring-Verfahren verwenden und auch die dazu definierten Größen und Matrizen. Hierzu ist es notwendig, entsprechend der penalisierten Loglikelihood für Zeitreihenbeobachtungen (4.4) den Loglikelihoodbeitrag für das Zeitintervall t , $l_t(\alpha_t)$, explizit anzugeben. Nach (4.35) ist er gegeben durch

$$l_t(\alpha_t) := \sum_{U \in R_t} \ln p(\mathbf{y}_{tU} | \mathbf{y}_{t-1}^*, \mathbf{x}_t^*, \mathbf{r}_t^*, \alpha_t) \quad (4.42)$$

die Summe der individuellen Loglikelihoodbeiträge der Untersuchungseinheiten U , die in t noch unter Risiko stehen. Dementsprechend fassen wir die individuellen Score-Funktionen und erwarteten Informationsmatrizen zusammen zur Score-Funktion von $l_t(\alpha_t)$ für das Zeitintervall t ,

$$\begin{aligned} \mathbb{R}^n \ni s_t(\alpha_t) &:= \partial l_t(\alpha_t) / \partial \alpha_t = \\ &= \begin{cases} \sum_{U \in R_t} \mathbf{Z}'_{tU} \mathbf{D}_{tU}(\alpha_t) \Sigma_{tU}^{-1}(\alpha_t) \{\mathbf{y}_{tU} - \lambda_{tU}(\alpha_t)\}, & \text{für das multinomiale Zustandsraummodell} \\ \sum_{U \in R_t} \mathbf{z}'_U (\mathbf{y}_{tU} - \mu_{tU}(\alpha_t)), \mu_{tU}(\alpha_t) = \Delta_{tU} \cdot \lambda_{tU}(\alpha_t), & \text{für das piecewise exponential Modell} \end{cases} \end{aligned} \quad (4.43)$$

und zur erwarteten Fisher-Informationsmatrix

$$\mathbb{R}^{n,n} \ni S_t(\alpha_t) := E\{-\partial^2 l_t(\alpha_t) / \partial \alpha_t \partial \alpha_t'\} =$$

$$= \begin{cases} \sum_{U \in R_t} Z'_{tU} D_{tU}(\alpha_t) \Sigma_{tU}^{-1}(\alpha_t) D'_{tU}(\alpha_t) Z_{tU}, & \text{für das multinomiale Zustandsraummodell} \\ \sum_{U \in R_t} \Delta_{tU} Z'_U \lambda_{tU}(\alpha_t) Z_U, & \text{für das piecewise exponential Modell} \end{cases}, \quad (4.44)$$

wobei für das multinomiale Zustandsraummodell gilt: $\lambda_{tU}(\alpha_t) = \mathbf{h}(\eta_{tU}) = \{\mathbf{h}_1(\eta_{tU}), \dots, \mathbf{h}_J(\eta_{tU})\}'$ sind die individuellen Hazard-Funktionen, $D_{tU}(\alpha_t) = \partial \mathbf{h}(\eta_{tU}) / \partial \eta$ die Jacobi-Matrix von $\mathbf{h}(\eta) = \{\mathbf{h}_1(\eta), \dots, \mathbf{h}_J(\eta)\}'$ ausgewertet an η_{tU} , und $\Sigma_{tU}(\alpha_t)$ die Kovarianzmatrix der Multinomialverteilung. Die Inverse $\Sigma_{tU}^{-1}(\alpha_t) \in \mathbb{R}^{\bar{J}, \bar{J}}$ dieser Kovarianzmatrix kann explizit angegeben werden:

$$\Sigma_{tU}^{-1}(\alpha_t) = \frac{1}{\sum_{j=1}^{\bar{J}} \mathbf{h}_j(\eta_{tU})} \begin{pmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{pmatrix} + \text{diag} \left\{ \frac{1}{\mathbf{h}_1(\eta_{tU})}, \dots, \frac{1}{\mathbf{h}_J(\eta_{tU})} \right\}, \quad (4.45)$$

siehe dazu auch Kredler (1986). Je nach Wahl von $\mathbf{h}(\eta_{tU})$ erhält man verschiedene Modelle. Für das dynamische multinomiale Logit-Modell (3.14) ist das Element in der i -ten Zeile und der j -ten Spalte, $i, j = 1, \dots, \bar{J}$, der Jacobi-Matrix $D_{tU}(\alpha_t)$ gegeben durch

$$\frac{\partial \mathbf{h}_j(\eta_{tU})}{\partial \eta_{tUi}} = \mathbf{h}_j(\eta_{tU}) \cdot \{\delta_{i,j} - \mathbf{h}_i(\eta_{tU})\}, \quad (4.46)$$

wobei

$$\delta_{i,j} = \begin{cases} 1, & \text{für } i = j \\ 0, & \text{sonst} \end{cases} \quad (4.47)$$

das Kronecker-Symbol ist. Wählt man die Response-Funktion wie in (3.15), so ergibt sich

$$\frac{\partial \mathbf{h}_j(\eta_{tU})}{\partial \eta_{tUj}} = \mathbf{h}_j(\eta_{tU}) + \frac{\exp(\eta_{tUj})}{\sum_{i=1}^{\bar{J}} \exp(\eta_{tUi})}.$$

$$\left[-\frac{\exp(\eta_{tUj})}{\sum_{i=1}^J \exp(\eta_{tUi})} + \frac{\exp\left\{-\sum_{i=1}^J \exp(\eta_{tUi})\right\}}{\sum_{i=1}^J \exp(\eta_{tUi})} + \exp(\eta_{tUj}) \cdot \exp\left\{-\sum_{i=1}^J \exp(\eta_{tUi})\right\} \right]$$

und für $j \neq i$

$$\frac{\partial h_j(\eta_{tU})}{\partial \eta_{tUi}} = \frac{\exp(\eta_{tUi})}{\sum_{s=1}^J \exp(\eta_{tUs})} \cdot \left[\exp(\eta_{tUj}) \cdot \exp\left\{-\sum_{s=1}^J \exp(\eta_{tUs})\right\} - h_j(\eta_{tU}) \right]. \quad (4.48)$$

Damit wir unsere penalisierte Loglikelihood wie im Zeitreihenfall in kompakter Matrixnotation angeben können, definieren wir

$$l(\alpha_{\bar{T}}^*) := \sum_{t=1}^{\bar{T}} l_t(\alpha_t) + l_0(\alpha_0) \quad (4.49)$$

mit $l_0(\alpha_0)$ wie in (4.8) und $l_t(\alpha_t)$ aus (4.42). Die penalisierte Loglikelihood (4.35) hat dann die gleiche Gestalt wie im Zeitreihenfall:

$$PL(\alpha_{\bar{T}}^*) = l(\alpha_{\bar{T}}^*) - \frac{1}{2} (\alpha_{\bar{T}}^*)' P(\alpha_{\bar{T}}^*), \quad (4.50)$$

mit P wie in (4.9). Um auch die Score-Funktion

$$s(\alpha_{\bar{T}}^*) = \partial l(\alpha_{\bar{T}}^*) / \partial \alpha_{\bar{T}}^* = \{s'_0(\alpha_0), s'_1(\alpha_1), \dots, s'_T(\alpha_T)\}'$$

und die erwartete Fisher-Informationsmatrix

$$S(\alpha_{\bar{T}}^*) = E\{-\partial^2 l(\alpha_{\bar{T}}^*) / \partial(\alpha_{\bar{T}}^*) \partial(\alpha_{\bar{T}}^*)'\} = \text{diag}\{S_0, S_1(\alpha_1), \dots, S_T(\alpha_T)\}$$

analog zum Zeitreihenfall in Matrixform darzustellen, definieren wir noch die folgenden

Größen für $t = 1, \dots, \bar{T}$: Die individuellen Hazardraten fassen wir in

$$\mu_t(\alpha_t) = \lambda_t(\alpha_t) := \{\lambda_{tU}(\alpha_t), U \in \mathbf{R}_t\} \quad (4.51)$$

zusammen, die individuellen Varianz-Kovarianzmatrizen in

$$\Sigma_t^{-1}(\alpha_t) := \text{diag} \left\{ \Sigma_{tU}^{-1}(\alpha_t), U \in \mathbf{R}_t \right\}, \quad (4.52)$$

die individuellen Jacobi-Matrizen in

$$D_t(\alpha_t) = \text{diag} \left\{ D_{tU}(\alpha_t), U \in \mathbf{R}_t \right\} \quad (4.53)$$

und die individuellen Design-Matrizen in der Matrix

$$Z_t = \begin{pmatrix} Z_{tU_1} \\ \vdots \\ Z_{tU_{|\mathbf{R}_t|}} \end{pmatrix}, \quad (4.54)$$

wobei Z_t all die individuellen Design-Matrizen Z_{tU_i} enthalten soll, für die gilt: $U_i \in \mathbf{R}_t$. Mit diesen Größen definieren wir $\mu(\alpha_{\bar{T}}^*)$ wie in (4.11), $D(\alpha_{\bar{T}}^*)$ wie in (4.14), Z wie in (4.13) und $\Sigma^{-1}(\alpha_{\bar{T}}^*) := \text{diag} \{ Q_0^{-1}, \Sigma_1^{-1}(\alpha_1), \dots, \Sigma_{\bar{T}}^{-1}(\alpha_{\bar{T}}) \}$. Mit y wie in (4.10), $s_0(\alpha_0)$ wie in (4.15) und S_0 wie in (4.17) lassen sich nun die Score-Funktion $s(\alpha_{\bar{T}}^*)$ und die erwartete Informationsmatrix $S(\alpha_{\bar{T}}^*)$ von $l(\alpha_{\bar{T}}^*)$ schreiben wie in (4.16) beziehungsweise (4.20). Die erste Ableitung $g(\alpha_{\bar{T}}^*)$ und die erwartete Informationsmatrix $G(\alpha_{\bar{T}}^*)$ der penalisierten Loglikelihood $PL(\alpha_{\bar{T}}^*)$ sind durch (4.21) und (4.22) gegeben. Insgesamt läßt sich ein Fisher-Scoring-Schritt von $\mathbf{a}^{(i)} \in \mathbb{R}^m$ nach $\mathbf{a}^{(i+1)} \in \mathbb{R}^m$ wie im Zeitreihenfall angeben:

$$\mathbf{a}^{(i+1)} = \{ Z'W(\mathbf{a}^{(i)})Z + P \}^{-1} Z'W(\mathbf{a}^{(i)})\tilde{y}(\mathbf{a}^{(i)}) \quad (4.55)$$

mit der „working observation“ $\tilde{y}(\mathbf{a}^{(i)})$ wie in (4.25). Das lineare Gleichungssystem (4.55) hat dieselbe Gestalt wie (4.24) im Zeitreihenfall. Wir können also zur Lösung das gleiche Verfahren anwenden. Man beachte jedoch, daß die Vektoren und Matrizen $\mu_t(\alpha_t)$, $\Sigma_t^{-1}(\alpha_t)$, $D_t(\alpha_t)$ und Z_t für Ereignisdaten im allgemeinen anders belegt sind. Ebenso bestehen die Score-Funktion und die erwartete Fisher-Informationsmatrix aus Summen von individuellen Beiträgen.

Für interne, zeitabhängige Kovariable betrachten wir wie in Abschnitt 3.2 den Loglikelihoodbeitrag $l_t(\alpha_t)$ in (4.42) für das Zeitintervall t als logarithmierten Partial-

Likelihood-Beitrag und $\sum_{t=1}^{\bar{T}} l_t(\alpha_t)$ aus (4.49) als Partial-Likelihood.

4.2 Mehr-Episoden-Fall

Im folgenden betrachten wir die Posteriori-Modus-Schätzung für den Mehr-Episoden-Fall. Die Kovariablen seien zunächst wieder extern.

Mit \mathbf{H}_{k-1} wie in (2.13) und den Werten der Kovariablen für die Untersuchungseinheit U in der k -ten Episode, $\mathbf{v}_U^k := (\mathbf{x}_{t_U^{k-1}, U}^k, \dots, \mathbf{x}_{t_U^k, U}^k)$, leitet Hamerle (1985b) unter den Annahmen $p(\mathbf{v}_U^k | \mathbf{H}_{k-1}, \mathbf{x}_{t_{k-1}}^{(k-1)*})$ unabhängig von α_t^k , $t=1, \dots, \bar{T}$, „random censoring“ und „nichtinformative“ Zensierung die folgende Formel für die gewöhnliche Likelihood L_G her:

$$L_G = \prod_{k=1}^{\bar{K}} \prod_{U=1}^{\bar{U}} \left[\prod_{j=1}^{\bar{J}} \left\{ \lambda_j^k \left(\mathbf{t}_U^k | \mathbf{x}_{(t_U^k)^*}^{k*} \right) \right\}^{y_{t_U^k, j}^k} \right] \cdot \left[\prod_{s=t_U^{k-1}}^{t_U^k-1} \left\{ 1 - \sum_{j=1}^{\bar{J}} \lambda_j^k (s | \mathbf{x}_s^{k*}) \right\} \right]^{t_{sU}^k}. \quad (4.56)$$

Wir nehmen an, daß die Hazardraten λ_j^k nur von den Zuständen α_t^k , $t=1, \dots, \bar{T}$, abhängen und nicht von α_t^l , $t=1, \dots, \bar{T}$, mit $l \neq k$. Damit muß bei einer Maximierung von L_G bezüglich α_t^k , $t=1, \dots, \bar{T}$, nur der Faktor für die k -te Episode berücksichtigt werden.

Umordnen in (4.56) in bezug auf die Zeit t und die Untersuchungseinheiten U ergibt

$$L_G = \prod_{k=1}^{\bar{K}} L_{G,k}(\alpha_{\bar{T}^*}^k) \text{ mit} \\ L_{G,k}(\alpha_{\bar{T}^*}^k) := \prod_{t=1}^{\bar{T}} \prod_{U \in R_t^k} \left[\prod_{j=1}^{\bar{J}} \left\{ \lambda_j^k (t | \mathbf{x}_t^{k*}) \right\}^{y_{tU,j}^k} \right] \cdot \left\{ 1 - \sum_{j=1}^{\bar{J}} \lambda_j^k (t | \mathbf{x}_t^{k*}) \right\}^{1 - \sum_{j=1}^{\bar{J}} y_{tU,j}^k}, \quad (4.57)$$

wobei $\alpha_{t^*}^k := (\alpha_0^k, \alpha_1^k, \dots, \alpha_t^k) \in \mathbb{R}^{n(t+1)}$, $t=1, \dots, \bar{T}$. $L_{G,k}(\alpha_{\bar{T}^*}^k)$ hat die Form einer Multinomiallikelihood im Ein-Episoden-Fall. Wenn wir für unser multinomiales Zustandsraummodell (3.12), (3.13) aus Abschnitt 3.1 die nachfolgenden Annahmen treffen, so erhalten wir die gleiche Gesamtlikelihood L_G wie in (4.57). Zunächst setzen wir wie in (4.30), jedoch auf den Mehr-Episoden-Fall verallgemeinert, voraus, daß

$$p(\mathbf{y}_{tU}^k | \mathbf{y}_{(t-1)^*}^{k*}, \mathbf{y}_t^{(k-1)*}, \mathbf{x}_{t^*}^{k*}, \mathbf{r}_{t^*}^{k*}, \alpha_t^k) = p(\mathbf{y}_{tU}^k | \mathbf{y}_{(t-1)^*}^{k*}, \mathbf{y}_t^{(k-1)*}, \mathbf{x}_{t^*}^{k*}, \mathbf{r}_{t^*}^{k*}), \quad \forall U \notin R_t^k, \quad t=1, \dots, \bar{T}. \quad (4.58)$$

Analog zu der Situation bei Querschnittsdaten nehmen wir auch an, daß

$$p(y_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_t^k) = \prod_{U=1}^{\bar{U}} p(y_{tU}^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_t^k), \quad t=1, \dots, \bar{T}. \quad (4.59)$$

Mit der gleichen Überlegung wie in (4.32) erhalten wir mit (4.58) und (4.59) insgesamt

$$p(y_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_t^k) \propto \prod_{U \in R_t^k} p(y_{tU}^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_t^k), \quad t=1, \dots, \bar{T}. \quad (4.60)$$

Zur Vereinfachung der Notation fassen wir in

$$\alpha_t^{k*} := (\alpha_t^l, l=1, \dots, k) \quad (4.61)$$

die Zustände im Zeitintervall t bis zur k-ten Episode und in

$$\alpha_{t^*}^{k*} := (\alpha_0^{k*}, \dots, \alpha_t^{k*}) = (\alpha_{t^*}^1, \alpha_{t^*}^2, \dots, \alpha_{t^*}^k) \quad (4.62)$$

alle Zustandsvektoren bis zur Zeit t und Episode Nummer k zusammen. Außerdem benötigen wir noch entsprechend (4.33) die nachstehenden drei Annahmen:

$$\begin{aligned} p(y_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_{t^*}^{k*}) &= p(y_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_t^k), \\ p(x_t^k, r_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{(t-1)}^{k*}, x_t^{(k-1)*}, r_{(t-1)}^{k*}, r_t^{(k-1)*}, \alpha_{(t-1)}^{k*}, \alpha_t^{(k-1)*}) &= \\ &= p(x_t^k, r_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{(t-1)}^{k*}, x_t^{(k-1)*}, r_{(t-1)}^{k*}, r_t^{(k-1)*}), \\ p(\alpha_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_{(t-1)}^{k*}, \alpha_t^{(k-1)*}) &= p(\alpha_t^k | \alpha_{t-1}^k), \quad t=1, \dots, \bar{T}. \end{aligned} \quad (4.63)$$

Damit gilt für die Posteriori-Dichte im Mehr-Zustand-Fall:

$$\begin{aligned} p(\alpha_{\bar{T}^*}^{k*} | y_{\bar{T}^*}^{k*}, x_{\bar{T}^*}^{k*}, r_{\bar{T}^*}^{k*}) &= \frac{1}{p(y_{\bar{T}^*}^{k*}, x_{\bar{T}^*}^{k*}, r_{\bar{T}^*}^{k*})} \cdot \\ &\cdot \left\{ \prod_{k=1}^{\bar{K}} \prod_{t=1}^{\bar{T}} p(y_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_{t^*}^{k*}) \cdot \right. \\ &\cdot \prod_{k=1}^{\bar{K}} \prod_{t=1}^{\bar{T}} p(x_t^k, r_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{(t-1)}^{k*}, x_t^{(k-1)*}, r_{(t-1)}^{k*}, r_t^{(k-1)*}, \alpha_{(t-1)}^{k*}, \alpha_t^{(k-1)*}) \cdot \\ &\cdot \left. \prod_{k=1}^{\bar{K}} \prod_{t=1}^{\bar{T}} p(\alpha_t^k | y_{(t-1)}^{k*}, y_t^{(k-1)*}, x_{t^*}^{k*}, r_{t^*}^{k*}, \alpha_{(t-1)}^{k*}, \alpha_t^{(k-1)*}) \cdot \prod_{k=1}^{\bar{K}} p(\alpha_0^k) \right\} \stackrel{(4.60), (4.63)}{\propto} \end{aligned}$$

$$\propto \prod_{k=1}^{\bar{K}} \left\{ \prod_{t=1}^{\bar{T}} \prod_{U \in \mathbb{R}_t^k} p(y_{tU}^k | y_{(t-1)*}^{k*}, y_t^{(k-1)*}, \mathbf{x}_{t*}^{k*}, \mathbf{r}_{t*}^{k*}, \alpha_t^k) \cdot \prod_{t=1}^{\bar{T}} p(\alpha_t^k | \alpha_{t-1}^k) \cdot p(\alpha_0^k) \right\}. \quad (4.64)$$

Logarithmieren auf der rechten Seite von (4.64) und Einsetzen der Normalverteilungsdichte aus dem Übergangsmodell (3.13) ergibt die penalisierte Loglikelihood für den Mehr-Episoden-Fall zu

$$\text{PL}(\alpha_{\bar{T}*}^{\bar{K}*}) = \sum_{k=1}^{\bar{K}} \text{PL}_k(\alpha_{\bar{T}*}^k), \quad (4.65)$$

wobei die penalisierte Loglikelihood $\text{PL}_k(\alpha_{\bar{T}*}^k)$ der k-ten Episode gegeben ist durch

$$\begin{aligned} \text{PL}_k(\alpha_{\bar{T}*}^k) &:= \sum_{t=1}^{\bar{T}} \sum_{U \in \mathbb{R}_t^k} p(y_{tU}^k | y_{(t-1)*}^{k*}, y_t^{(k-1)*}, \mathbf{x}_{t*}^{k*}, \mathbf{r}_{t*}^{k*}, \alpha_t^k) - \frac{1}{2} (\alpha_0^k - \mathbf{a}_0^k)' (\mathbf{Q}_0^k)^{-1} (\alpha_0^k - \mathbf{a}_0^k) \\ &\quad - \frac{1}{2} \sum_{t=1}^{\bar{T}} (\alpha_t^k - \mathbf{F}_t^k \alpha_{t-1}^k)' (\mathbf{Q}_t^k)^{-1} (\alpha_t^k - \mathbf{F}_t^k \alpha_{t-1}^k). \end{aligned} \quad (4.66)$$

Aus (4.65) wird deutlich, daß man $\text{PL}(\alpha_{\bar{T}*}^{\bar{K}*})$ getrennt für jede Episode k maximieren kann.

Das bedeutet, der Posteriori-Modus $\mathbf{a} \in \mathbb{R}^{\mathbf{m} \cdot \bar{K}}$

$$\mathbf{a} = (\mathbf{a}'_1, \mathbf{a}'_2, \dots, \mathbf{a}'_{\bar{K}})' := \arg \max_{\alpha_{\bar{T}*}^{\bar{K}*}} \{p(\alpha_{\bar{T}*}^{\bar{K}*} | \mathbf{y}_{\bar{T}*}^{\bar{K}*}, \mathbf{x}_{\bar{T}*}^{\bar{K}*}, \mathbf{r}_{\bar{T}*}^{\bar{K}*})\} \quad (4.67)$$

im Mehr-Episoden-Fall läßt sich dadurch berechnen, daß man für jede Episode $k = 1, \dots, \bar{K}$ die penalisierten Loglikelihood $\text{PL}_k(\alpha_{\bar{T}*}^k)$ der k-ten Episode aus (4.66) maximiert:

$$\mathbf{a}_k = \arg \max_{\alpha_{\bar{T}*}^k} \{\text{PL}_k(\alpha_{\bar{T}*}^k)\}. \quad (4.68)$$

Dieses Optimierungsproblem wiederum hat die gleiche Struktur wie die Maximierung der penalisierten Likelihood $\text{PL}(\alpha_{\bar{T}}^*)$ aus (4.35) im Ein-Episoden-Fall. Somit läßt sich die Bestimmung des Posteriori-Modus im Mehr-Episoden-Fall zurückführen auf den Ein-Episoden-Fall. Verwendet man dazu die Fisher-Scoring-Methode, so sind lineare Gleichungssysteme der Form (4.55) zu lösen. Diese wiederum besitzen die gleiche Struktur wie die linearen Gleichungssysteme (4.24) im Zeitreihenfall, mit entsprechend spezifizierten Score-Funktionen $s_t(\alpha_t)$ aus (4.15) beziehungsweise (4.43) und Fisher-Informationsmatrizen $\mathbf{S}_t(\alpha_t)$ aus (4.17) respektive (4.44). Insgesamt gilt: Hat man ein effizientes Verfahren zur Lösung von linearen Gleichungssystemen der Form (4.24), so kann damit

über die Fisher-Scoring-Methode genauso effizient (4.68) und daher auch (4.67) gelöst werden. Ein solches Verfahren werden wir im nächsten Kapitel besprechen.

Verwenden wir im speziellen das multinomiale Zustandsraummodell (3.12), (3.13) aus Abschnitt 3.1, so ergibt sich für die penalisierten Loglikelihood $\mathbf{PL}_k(\alpha_{\bar{T}^*}^k)$ der k -ten Episode:

$$\begin{aligned} \mathbf{PL}_k(\alpha_{\bar{T}^*}^k) = & \ln\{\mathbf{L}_{G,k}(\alpha_{\bar{T}^*}^k)\} - \frac{1}{2}(\alpha_0^k - \mathbf{a}_0^k)' (\mathbf{Q}_0^k)^{-1}(\alpha_0^k - \mathbf{a}_0^k) \\ & - \frac{1}{2} \sum_{t=1}^{\bar{T}} (\alpha_t^k - \mathbf{F}_t^k \alpha_{t-1}^k)' (\mathbf{Q}_t^k)^{-1}(\alpha_t^k - \mathbf{F}_t^k \alpha_{t-1}^k), \end{aligned} \quad (4.69)$$

mit $\mathbf{L}_{G,k}(\alpha_{\bar{T}^*}^k)$ wie in (4.57).

In diesem Abschnitt wurde bisher für die Kovariablen \mathbf{x}_t angenommen, daß sie extern sind. Für interne, zeitabhängige Kovariablen zeigt Hamerle (1985b), daß man die gewöhnliche Likelihood \mathbf{L}_G bei externen Kovariablen in (4.56) und damit auch in (4.57) unter den gleichen Voraussetzungen wie bei (4.56) als Partial-Likelihood verwenden kann, denn die gewöhnliche Likelihood $\tilde{\mathbf{L}}_G$ für interne Kovariable ist gegeben durch

$$\tilde{\mathbf{L}}_G = \prod_{k=1}^{\bar{K}} \mathbf{L}_{G,k}(\alpha_{\bar{T}^*}^k) \cdot \mathbf{Q}_k. \quad (4.70)$$

Der Faktor \mathbf{Q}_k ist in Hamerle (1985b, Formel 4.2) angegeben. Ist \mathbf{Q}_k unabhängig von den Regressionskoeffizienten $\alpha_{\bar{T}^*}^{\bar{K}^*}$, dann gilt: $\tilde{\mathbf{L}}_G \propto \mathbf{L}_G$. In diesem Sinne wollen wir auch die penalisierte Loglikelihood für den Mehr-Episoden-Fall, $\mathbf{PL}(\alpha_{\bar{T}^*}^{\bar{K}^*})$, bei Vorliegen von internen Kovariablen als Penalized-Partial-Loglikelihood verwenden.

5 Schätzalgorithmen

In diesem Kapitel werden wir Algorithmen zur Lösung des linearen Gleichungssystems (4.24) vorstellen und beschreiben. Nach unseren Ausführungen am Ende von Abschnitt 4.2 liegen damit auch Verfahren zur Bestimmung des Posteriori-Modus im Ein- und Mehr-Zustands-Fall vor.

Zunächst behandeln wir den Fall, daß eine informative Start-Priori-Verteilung für den Anfangszustand α_0 vorliegt. Anschließend gehen wir auf die Situation einer diffusen, nicht-informativen Start-Priori ein. Am Ende dieses Kapitels stellen wir einige numerische Aspekte dar, die im Zusammenhang mit der Implementierung der besprochenen Schätzalgorithmen relevant sind, wie zum Beispiel die Vermeidung von Matrixinvertierungen.

5.1 Schätzalgorithmen für Modelle mit informativer Start-Priori

Eine informative Start-Priori-Verteilung für den Anfangszustand α_0 im Übergangsmodell (3.7) liegt vor, wenn „die Varianz von α_0 endlich ist“, genauer: $\text{cov}(\alpha_0) = Q_0 \in \mathbb{R}^{n,n}$ und Q_0 positiv semidefinit. Der Erwartungswertvektor a_0 von α_0 , die Kovarianzmatrix Q_0 und $\text{cov}(\xi_t) = Q_t$ werden als Hyperparameter bezeichnet. Für a_0 und Q_0 verwenden wir auch den Begriff Start-Hyperparameter.

Im folgenden nehmen wir an, daß die Hyperparameter unseres Zustandsraummodells fest vorgegeben sind. Lediglich die Zustände α_t sind die nicht beobachtbaren, zu schätzenden Größen. Die Situation, in der auch die Hyperparameter unbekannt und daher zu schätzen sind, wird in Teilabschnitt 5.1.4 behandelt.

Für den Fall des linearen Zustandsraummodells (3.1), (3.2) haben wir in Kapitel 4 bereits angedeutet, daß das lineare Gleichungssystem (4.26) mit dem (KFG)-Algorithmus gelöst werden kann. Darauf wollen wir im folgenden näher eingehen und den Kalman Filter und Glätter vorstellen. Wegen der Normalverteilungsannahme in (3.1) und (3.2) ist die Posteriori-Dichte ebenfalls die Normalverteilungsdichte mit den Erwartungswerten

$$a_{t|\bar{T}} := E(\alpha_t | y_{\bar{T}}^*, x_{\bar{T}}^*) \quad (5.1)$$

und den Kovarianzen

$$\mathbf{V}_{t|\bar{T}} := \text{cov}(\alpha_t | \mathbf{y}_{\bar{T}}^*, \mathbf{x}_{\bar{T}}^*) , \quad t = 0, \dots, \bar{T} . \quad (5.2)$$

Die Größen in (5.1) und (5.2) werden im Kalman-Filter-Kontext auch als Glättungslösung bezeichnet. Von Filterwerten spricht man bei

$$\begin{aligned} \mathbf{a}_{t|t} &:= E(\alpha_t | \mathbf{y}_t^*, \mathbf{x}_t^*) \quad \text{und} \\ \mathbf{V}_{t|t} &:= \text{cov}(\alpha_t | \mathbf{y}_t^*, \mathbf{x}_t^*) , \end{aligned} \quad (5.3)$$

und von s-Schritt Prognosewerten ($s \in \mathbb{N}$) bei

$$\begin{aligned} \mathbf{a}_{t+s|t} &:= E(\alpha_{t+s} | \mathbf{y}_t^*, \mathbf{x}_t^*) \quad \text{und} \\ \mathbf{V}_{t+s|t} &:= \text{cov}(\alpha_{t+s} | \mathbf{y}_t^*, \mathbf{x}_t^*) , \quad t = 0, \dots, \bar{T} . \end{aligned} \quad (5.4)$$

Anderson und Moore (1979) zeigen, daß sich die Erwartungswerte und Kovarianzmatrizen in (5.1) – (5.4) auf die folgende, rekursive Weise berechnen lassen:

Kalman Filter und Glätter (KFG):

Initialisierung: $\mathbf{a}_{0|0} = \mathbf{a}_0 , \quad \mathbf{V}_{0|0} = \mathbf{Q}_0 .$

Für $t = 1, \dots, \bar{T}$:

$$\begin{aligned} \text{Vorhersageschritt:} \quad \mathbf{a}_{t|t-1} &= \mathbf{F}_t \mathbf{a}_{t-1|t-1} , \\ \mathbf{V}_{t|t-1} &= \mathbf{F}_t \mathbf{V}_{t-1|t-1} \mathbf{F}_t' + \mathbf{Q}_t . \end{aligned}$$

$$\begin{aligned} \text{Korrekturschritt:} \quad \mathbf{V}_{t|t} &= \left\{ \mathbf{V}_{t|t-1}^{-1} + \mathbf{Z}_t' \Sigma_t^{-1} \mathbf{Z}_t \right\}^{-1} , \\ \mathbf{a}_{t|t} &= \mathbf{a}_{t|t-1} + \mathbf{V}_{t|t} \mathbf{Z}_t' \Sigma_t^{-1} (\mathbf{y}_t - \mathbf{Z}_t \mathbf{a}_{t|t-1}) . \end{aligned}$$

Für $t = \bar{T}, \dots, 1$:

$$\begin{aligned} \text{Glättungsschritt:} \quad \mathbf{B}_t &= \mathbf{V}_{t-1|t-1} \mathbf{F}_t' \mathbf{V}_{t|t-1}^{-1} \\ \mathbf{a}_{t-1|\bar{T}} &= \mathbf{a}_{t-1|t-1} + \mathbf{B}_t (\mathbf{a}_{t|\bar{T}} - \mathbf{a}_{t|t-1}) , \\ \mathbf{V}_{t-1|\bar{T}} &= \mathbf{V}_{t-1|t-1} + \mathbf{B}_t (\mathbf{V}_{t|\bar{T}} - \mathbf{V}_{t|t-1}) \mathbf{B}_t' . \end{aligned}$$

Wir nehmen, daß sämtliche Matrizen, die in obigem Verfahren invertiert werden müssen, auch regulär sind. Der Glättungsschritt in (KFG) wird als „fixed-interval-Glätter“ bezeichnet. Weitere Möglichkeiten für die Glättung finden sich in Anderson und Moore (1979) und Koopman (1993).

Der Kalman Filter und Glätter liefert also für das lineare Zustandsraummodell (3.1), (3.2) den Posteriori-Erwartungswertvektor $\mathbf{a} = (\mathbf{a}_{0|\bar{T}}, \mathbf{a}_{1|\bar{T}}, \dots, \mathbf{a}_{\bar{T}|\bar{T}})$. Weil bei der Normalverteilung Erwartungswert und Modus übereinstimmen, löst der (KFG)-Algorithmus wegen (4.5) auch das quadratische Optimierungsproblem in (4.7), wobei $\mathbf{a} \in \mathbb{R}^m$ der Vektor ist, an dem die quadratische Funktion $PL(\alpha_{\bar{T}}^*)$ ihr Maximum annimmt. \mathbf{a} ist damit auch ein stationärer Punkt, das heißt ein Vektor, für den gilt:

$$\mathbf{g}(\mathbf{a}) = \frac{\partial PL(\mathbf{a})}{\partial \alpha_{\bar{T}}^*} = \mathbf{0}_m. \quad (5.5)$$

Dieses lineare Gleichungssystem läßt sich (Invertierbarkeit vorausgesetzt) äquivalent umformen zu dem dünnbesetzten linearen Gleichungssystem (4.26) mit Blocktridiagonalstruktur. Der (KFG)-Algorithmus löst also (4.26). Dies geschieht ohne explizite Invertierung der blocktridiagonalen Gleichungsmatrix $\mathbf{Z}'\Sigma^{-1}\mathbf{Z} + \mathbf{P}$. Außerdem nutzt das (KFG)-Verfahren die Blocktridiagonalgestalt aus und daher steigt der numerische Aufwand zur Berechnung der Lösung nur linear mit \bar{T} . Bei einer Invertierung wäre der numerische Aufwand weitaus größer, denn die Inverse einer tridiagonalen Matrix ist im allgemeinen vollbesetzt. Damit ist der Kalman Filter und Glätter aus numerischer Sicht ein effizientes Verfahren zur Lösung blocktridiagonaler linearer Gleichungssysteme der Form (4.26).

Unser eigentliches Ziel ist es jedoch, (4.24) zu lösen. Wie dies auf effiziente Weise geschieht und wie dazu der (KFG)-Algorithmus eingesetzt werden kann, zeigen wir im nächsten Teilabschnitt.

5.1.1 Linear gewichteter Kalman Filter und Glätter (LGKFG)

Vergleicht man die linearen Gleichungssysteme (4.24) und (4.26) miteinander, so ergibt sich das Folgende: Ersetzt man in (4.26) die Beobachtungen \mathbf{y}_t durch die „working observation“ $\tilde{\mathbf{y}}_t(\mathbf{a}_t^{(i)})$ aus (4.25) und die inverse Fehler-Kovarianzmatrix Σ_t^{-1} aus dem

Beobachtungsmodell (3.1) durch die Gewichtsmatrix $W_t(a_t^{(i)})$ für $t = 1, \dots, \bar{T}$ unter Beachtung der Definition von $\tilde{y}(a^{(i)})$ in (4.25) und von $W(a^{(i)})$ in (4.19), so erhält man (4.24). Weil der Kalman Filter und Glätter (4.26) löst, braucht man in (KFG) nur diese Ersetzungen vorzunehmen, um ein Lösungsverfahren für (4.24) zu bekommen. Den daraus resultierenden Algorithmus wollen wir bezeichnen als

Linear gewichteter Kalman Filter und Glätter (LGKFG):

Initialisierung: $a_{0|0} = a_0$, $V_{0|0} = Q_0$.

Für $t = 1, \dots, \bar{T}$:

Vorhersageschritt: Wie in (KFG).

Korrekturschritt:
$$V_{t|t} = \{V_{t|t-1}^{-1} + S_t(a_t^{(i)})\}^{-1},$$

$$\tilde{s}_t(a_t^{(i)}) = s_t(a_t^{(i)}) - S_t(a_t^{(i)})\{a_{t|t-1} - a_t^{(i)}\}$$

$$a_{t|t} = a_{t|t-1} + V_{t|t} \tilde{s}_t(a_t^{(i)}).$$

Falls $t = \bar{T}$, dann setze $a_{\bar{T}|\bar{T}}^{(i+1)} = a_{\bar{T}|\bar{T}}$.

Für $t = \bar{T}, \dots, 1$:

Glättungsschritt:
$$B_t = V_{t-1|t-1} F_t' V_{t|t-1}^{-1}$$

$$a_{t-1|\bar{T}}^{(i+1)} = a_{t-1|t-1} + B_t(a_{t|\bar{T}}^{(i+1)} - a_{t|t-1}),$$

$$V_{t-1|\bar{T}} = V_{t-1|t-1} + B_t(V_{t|\bar{T}} - V_{t|t-1})B_t'.$$

Die Lösung von (LGKFG) ist gegeben durch $a^{(i+1)} = ((a_{0|\bar{T}}^{(i+1)})', (a_{1|\bar{T}}^{(i+1)})', \dots, (a_{\bar{T}|\bar{T}}^{(i+1)})')'$.

$\tilde{s}_t(a_t^{(i)})$ im Korrekturschritt von (LGKFG) bezeichnen wir als „working score“ Funktion,

ausgewertet an der Stelle $\mathbf{a}_t^{(i)}$. Der (LGKFG)-Algorithmus zeichnet sich durch drei wesentliche Eigenschaften aus:

- (i) (LGKFG) löst das lineare Gleichungssystem (4.24) genauso effizient wie (KFG) die Gleichungen in (4.26). Insbesondere wird die explizite Invertierung der blocktri-diagonalen Gleichungsmatrix $\mathbf{Z}'\mathbf{W}(\mathbf{a}^{(i)})\mathbf{Z} + \mathbf{P}$ in (4.24) vermieden und ihre dünn-besetzte Struktur ausgenutzt.
- (ii) (LGKFG) führt einen Fisher-Scoring-Schritt von $\mathbf{a}^{(i)}$ nach $\mathbf{a}^{(i+1)}$ durch.
- (iii) Nach Fahrmeir und Kaufmann (1991, Proposition 1) sind die Matrizen $\mathbf{V}_{t|\bar{T}} \in \mathbb{R}^{n,n}$, $t = 0, \dots, \bar{T}$ die Hauptdiagonalblöcke von $\mathbf{G}(\mathbf{a}^{(i)})^{-1}$, der inversen Fisher-Informationsmatrix von $\text{PL}(\alpha_{\bar{T}}^*)$, ausgewertet am Iterationspunkt $\mathbf{a}^{(i)}$.

Um einen Fisher-Scoring-Schritt von $\mathbf{a}^{(i)}$ nach $\mathbf{a}^{(i+1)}$, $i = 1, 2, \dots$, mit (LGKFG) durchführen zu können, benötigt man einen Startvektor $\mathbf{a}^{(1)} \in \mathbb{R}^m$. Wie man auf günstige Weise zu einem Startvektor kommt, beschreiben wir im nächsten Teilabschnitt.

5.1.2 Verallgemeinerter und erweiterter Kalman Filter und Glätter (EKFG)

Fahrmeir (1992) schlägt zur „approximativen“ Lösung des Posteriori-Modus-Schätzproblems (4.5) den verallgemeinerten und erweiterter Kalman Filter und Glätter vor. Das Verfahren ist durch die folgenden Rekursionen gegeben:

Verallgemeinerter und erweiterter Kalman Filter und Glätter (EKFG):

Initialisierung: $\mathbf{a}_{0|0} = \mathbf{a}_0$, $\mathbf{V}_{0|0} = \mathbf{Q}_0$.

Für $t = 1, \dots, \bar{T}$:

Vorhersageschritt: Wie in (KFG).

$$\begin{aligned} \text{Korrekturschritt: } V_{t|t} &= \left\{ V_{t|t-1}^{-1} + S_t(a_{t|t-1}) \right\}^{-1}, \\ a_{t|t} &= a_{t|t-1} + V_{t|t} s_t(a_{t|t-1}). \end{aligned}$$

Falls $t = \bar{T}$, dann setze $a_{\bar{T}|\bar{T}}^{(i)} = a_{\bar{T}|\bar{T}}$.

Für $t = \bar{T}, \dots, 1$:

$$\begin{aligned} \text{Glättungsschritt: } B_t &= V_{t-1|t-1} F_t' V_{t|t-1}^{-1} \\ a_{t-1|\bar{T}}^{(i)} &= a_{t-1|t-1} + B_t (a_{t|\bar{T}}^{(i)} - a_{t|t-1}), \\ V_{t-1|\bar{T}} &= V_{t-1|t-1} + B_t (V_{t|\bar{T}} - V_{t|t-1}) B_t'. \end{aligned}$$

Wir haben den Korrekturschritt wie auch bei (KFG) und (LGKFG) in der (globalen) Scoring-Form angegeben. Mit einem Matrix-Inversionslemma (vergleiche etwa Anderson und Moore, 1979, Abschnitt 6.3) ergibt sich daraus sofort die bekannte Kalman-Gain-Form. Als Lösung von (EKFG) erhalten wir nach Fahrmeir (1992) einen approximativen Posteriori-Modus-Schätzer, den wir mit $a^{(i)} = ((a_0^{(i)})', (a_1^{(i)})', \dots, (a_{\bar{T}}^{(i)})')' \in \mathbb{R}^m$ bezeichnen wollen.

Eine andere, direkte Ableitung von (EKFG) kann man aus dem (LGKFG)-Verfahren gewinnen: Setzt man für den Eingangsvektor $a^{(i)}$ von (LGKFG)

$$a^{(i)} = (a'_0, a'_{1|0}, a'_{2|1}, \dots, a'_{t|t-1}, \dots, a'_{\bar{T}|\bar{T}-1})', \quad (5.6)$$

so erhält man die (EKFG)-Formeln mit $a^{(i)} = a^{(i+1)}$. Damit läßt sich der (EKFG) auch als Spezialfall von (LGKFG) auffassen. Die einzelnen Komponenten von $a^{(i)}$ auf der rechten Seite in (5.6) sind entweder durch das jeweilige Zustandsraummodell vorgegeben oder werden im Laufe der (EKFG)-Rekursionen berechnet. In diesem Sinne wählt sich das (EKFG)-Verfahren den Eingangsvektor $a^{(i)}$ „selbst“. Er muß also nicht wie bei (LGKFG) „von außen“ durch den Benutzer vorgegeben werden. Damit eignet sich der (EKFG) auf ideale Weise als Startverfahren zur Gewinnung eines Eingangsvektors $a^{(i)}$ für das (LGKFG)-Verfahren.

5.1.3 Iterativ gewichteter Kalman Filter und Glätter (IGKFG)

Bis jetzt haben wir uns mit der Durchführung eines Fisher-Scoring-Schrittes beschäftigt. Für das Fisher-Scoring-Verfahren selbst müssen wir den (LGKFG)-Algorithmus iterieren, wobei die Lösung $\mathbf{a}^{(i)}$ der vorangegangenen (LGKFG)-Rekursionen als Eingangsvektor für die nachfolgenden (LGKFG)-Rekursionen zu verwenden ist. Den (EKFG) benutzen wir als Startverfahren. Insgesamt erhalten wir damit einen iterativen Algorithmus, den wir bezeichnen als

Iterativ gewichteter Kalman Filter und Glätter (IGKFG):

Initialisierung: Berechne $\mathbf{a}^{(1)} = ((\mathbf{a}_0^{(1)})', (\mathbf{a}_1^{(1)})', \dots, (\mathbf{a}_T^{(1)})')' \in \mathbb{R}^m$ mit (EKFG).

Setze den Iterationsindex $i = 1$.

Schritt 1: Mit dem Eingangsvektor $\mathbf{a}^{(i)} \in \mathbb{R}^m$ berechne $\mathbf{a}^{(i+1)} \in \mathbb{R}^m$ unter Verwendung von (LGKFG).

Schritt 2: Wenn ein Abbruchkriterium erfüllt ist: STOP.

Anderenfalls erhöhe den Iterationsindex i um 1 und gehe zu Schritt 1.

Für die Wahl eines Abbruchkriteriums siehe beispielsweise Gill, Murray und Wright (1981). Als Lösung von (IGKFG) erhalten wir den Posteriori-Modus-Schätzer $\mathbf{a} \in \mathbb{R}^m$ aus (4.2) beziehungsweise (4.5). Zwei Eigenschaften von (IGKFG) sollen im folgenden hervorgehoben werden:

- (i) (IGKFG) ist ein Fisher-Scoring-Verfahren zur Lösung von nichtlinearen Optimierungsproblemen der Form (4.5). Die einzelnen Fisher-Scoring-Iterationen werden auf effiziente Weise durchgeführt.

(ii) (IGKFG) liefert automatisch und ohne zusätzlichen numerischen Aufwand mit den Matrizen $V_{t|\bar{T}}$, $t = 0, 1, \dots, \bar{T}$, die Hauptdiagonalblöcke $G(\mathbf{a}^{(i)})^{-1}$ der inversen Fisher-Informationsmatrix von $PL(\alpha_{\bar{T}}^*)$, ausgewertet am Iterationspunkt $\mathbf{a}^{(i)}$. Im Konvergenzfall erhält man also mit $V_{t|\bar{T}}$ die Hauptdiagonalblöcke von $G(\mathbf{a})^{-1}$ an der Stelle des Posteriori-Modus-Schätzers $\mathbf{a} \in \mathbb{R}^m$.

Die Hyperparameter waren bisher als fest vorgegeben angenommen worden. Außer bei simulierten Daten wird man auf diese Situation in der Praxis selten treffen. Aus diesem Grunde wollen wir im nachfolgenden Teilabschnitt eine Methode angeben, wie die Hyperparameter aus den Daten geschätzt werden können.

5.1.4 EM-Typ Algorithmus für informative Start-Priori

Zur datengesteuerten Hyperparameterschätzung gibt es verschiedene Ansatzmöglichkeiten. Insbesondere bei semi- und nonparametrischen Modellen werden in der Literatur eine Vielzahl von Kriterien zur Glättungsparameterwahl besprochen (siehe etwa Härdle, 1993, und die darin enthaltenen Literaturhinweise). Wagenpfeil (1995) vergleicht im Rahmen von verallgemeinerten Zustandsraummodellen drei Methoden zur Hyperparameterwahl: zum einen die Idee der approximativen Likelihood, die von Durbin und Koopman (1992) vorgeschlagen wurde, zum anderen den in Fahrmeir (1994) besprochenen EM-Typ Algorithmus und zum dritten die Kreuzvalidierung. Mit den Hauptdiagonalblöcken der inversen Fisher-Informationsmatrix, die man aus der Anwendung von (IGKFG) erhält, hat man Zugang zur „Hat“-Matrix und damit auch zum verallgemeinerten Kreuzvalidierungskriterium (siehe dazu Fahrmeir und Wagenpfeil, 1995a, Abschnitt 3.2, und Wagenpfeil, 1995). Das Ergebnis dieses Vergleiches läßt sich folgendermaßen zusammenfassen: Der EM-Typ Algorithmus ist ein robustes Verfahren zur Hyperparameterschätzung, allerdings mit langsamer Konvergenz. Die approximative Likelihood und die Kreuzvalidierung sind numerisch anfälliger, aber schneller, wenn sie funktionieren.

Aufgrund seiner Robustheit wurde im Rahmen dieser Arbeit bei der Hyperparameterwahl stets der EM-Typ Algorithmus verwendet. Wir werden das Verfahren im folgenden darstellen und in Teilabschnitt 5.2.4 für die Situation einer diffusen Start-Priori-Verteilung modifizieren. Um nicht eine zu große Anzahl von Parametern schätzen zu müssen, treffen

wir noch die Vereinbarung, daß Q_t nicht über die Zeit variieren soll, sondern $Q := Q_1 = Q_2 = \dots = Q_{\bar{T}}$.

EM-Typ Algorithmus:

Schritt 1: Wähle Startwerte $Q^{(0)}, Q_0^{(0)}, a_0^{(0)}$ und setze den Iterationsindex $p = 0$.

Schritt 2: Berechne $a_{t|\bar{T}}, V_{t|\bar{T}}$ ($t = 0, \dots, \bar{T}$) mit (EKFG) oder (IGKFG), unter Verwendung von $Q^{(p)}, Q_0^{(p)}$ und $a_0^{(p)}$.

Schritt 3: EM-Typ Schritt: Berechne $Q^{(p+1)}, Q_0^{(p+1)}$ und $a_0^{(p+1)}$ mit den zuvor in Schritt 2 bestimmten Größen $a_{t|\bar{T}}, V_{t|\bar{T}}$ ($t = 0, \dots, \bar{T}$) durch

$$a_0^{(p+1)} = a_{0|\bar{T}},$$

$$Q_0^{(p+1)} = V_{0|\bar{T}},$$

$$Q^{(p+1)} = \frac{1}{\bar{T}} \sum_{t=1}^{\bar{T}} \left\{ \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right) \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right)' + V_{t|\bar{T}} - F_t B_t V_{t|\bar{T}} - V_{t|\bar{T}}' B_t' F_t' + F_t V_{t-1|\bar{T}} F_t' \right\}$$

mit B_t wie im Glättungsschritt von (IGKFG) oder (EKFG).

Schritt 4: Wenn ein Abbruchkriterium erreicht wurde: STOP.

Sonst erhöhe den Iterationsindex p um 1 und gehe zu Schritt 2.

Als Lösung erhalten wir simultan die geschätzten Hyperparameter $\hat{a}_0, \hat{Q}_0, \hat{Q}$ und den Posteriori-Modus-Schätzer $a \in \mathbb{R}^m$, berechnet mit den Hyperparameterwerten $\hat{a}_0, \hat{Q}_0, \hat{Q}$. Der EM-Typ Algorithmus ist also ein Verfahren, bei dem die unbekannten Zustände α_t , $t = 0, \dots, \bar{T}$ und die Hyperparameter a_0, Q_0 und Q gemeinsam geschätzt werden. (IGKFG) oder (EKFG) wird dabei als Unterprogramm verwendet.

5.2 Schätzalgorithmen für Modelle mit diffuser Start-Priori

Eine diffuse Start-Priori-Verteilung liegt vor, wenn in unserem Übergangsmodell (3.2) und (3.7) $\text{cov}(\alpha_0) = Q_0$ „unendlich“ groß ist. Wir setzen dann $Q_0^{-1} = 0_{n,n}$.

Die Wahl einer diffusen Start-Priori-Verteilung kann aus zwei Gründen sinnvoll sein:

- (i) Die Start-Hyperparameter a_0 und Q_0 im Übergangsmodell (3.2) oder (3.7) sind nicht bekannt. Außer in Simulationsstudien ist das der Regelfall. Bei einer diffusen Start-Priori-Verteilung im Übergangsmodell sind dann nur noch die Fehler-Kovarianzmatrizen Q_t , $t = 1, \dots, \bar{T}$, unbekannte Hyperparameter.
- (ii) Es treten immer wieder Probleme bei der datengesteuerten Schätzung der Start-Hyperparameter auf. Anhand von zahlreichen Programmläufen läßt sich erkennen, daß beim EM-Typ Algorithmus Q_0 stark unterschätzt wird, möglicherweise bedingt durch die Festlegung $Q_1 = Q_2 = \dots = Q_{\bar{T}}$. Daraus resultiert eine auch in Aydemir (1992) festgestellte, sehr starke Startwertabhängigkeit der Schätzung \hat{a}_0 vom Startwert $a_0^{(0)}$. Außerdem sind die Konfidenzbänder für die Schätzungen von α_1 im Bereich um $t = 0$ „unnatürlich“ schmal. Ein Vergleich von Schätzergebnissen, die man einmal mit informativer, einmal mit diffuser Start-Priori-Verteilung erhalten hat, macht dies deutlich.

Die folgenden beiden Abbildungen beziehen sich auf die Efron-Daten aus Teilabschnitt 1.1.2 und zeigen den geschätzten Therapieeffekt einmal mit informativer Start-Priori (Abbildung 5.1) und einmal mit diffuser Start-Priori (Abbildung 5.2). Dabei wurde stets das dynamische binomiale Logit-Modell verwendet, das in Abschnitt 6.2 angegeben ist.

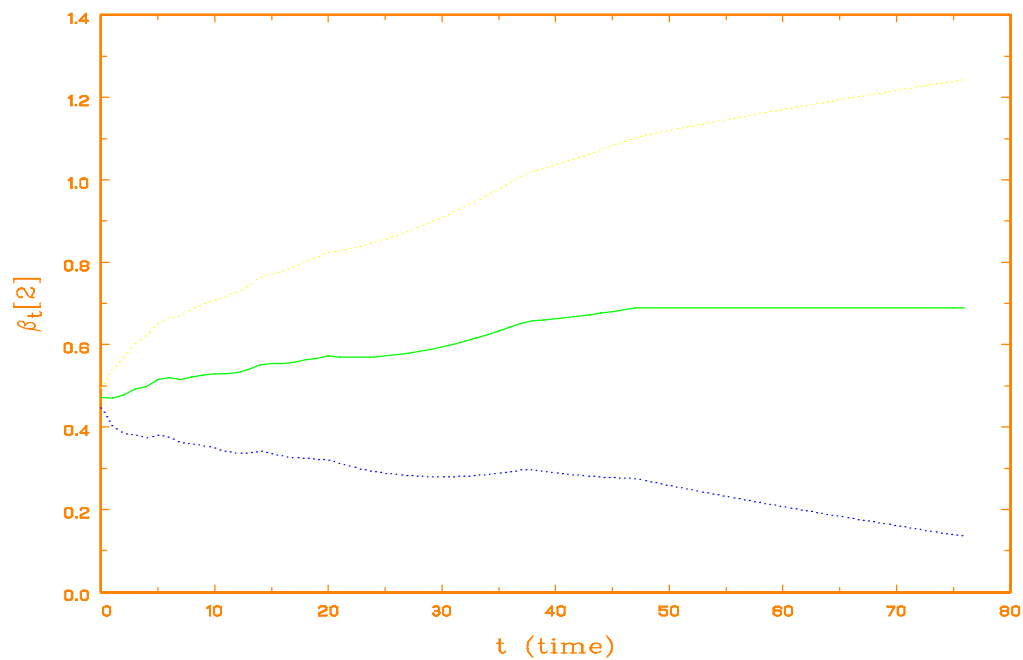


Abbildung 5.1 Efron-Daten. Geschätzter Therapieeffekt mit informativer Start-Priori.

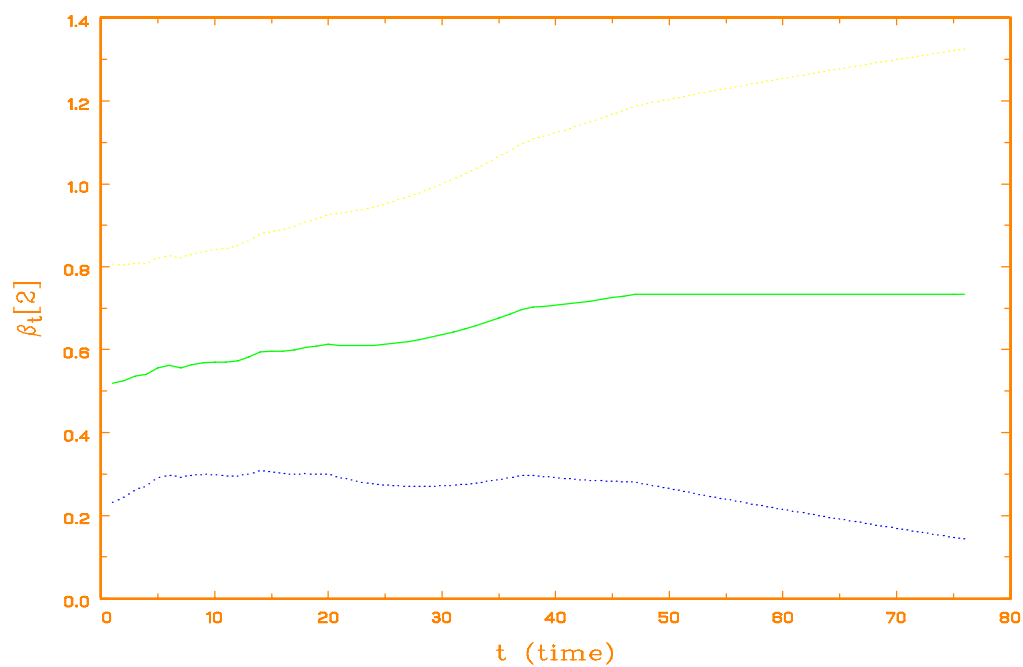


Abbildung 5.2 Efron-Daten. Geschätzter Therapieeffekt mit diffuser Start-Priori.

In Abbildung 5.1 ziehen sich die geschätzten Konfidenzbänder zum Zeitintervall Null hin „unnatürlich“ eng zusammen. Dies ist in Abbildung 5.2 nicht der Fall.

Um zu einem Schätzverfahren bei diffuser Start-Priori-Verteilung zu kommen gibt es prinzipiell zwei Möglichkeiten des Vorgehens:

- (i) Harvey (1992, Abschnitt 3.3) schlägt vor, im (KFG)-Algorithmus $V_{0|0} = Q_0 = l \cdot I_n$ zu wählen mit l eine sehr große, aber endliche Zahl.
- (ii) Die analytische Ableitung der (KFG)-Rekursionen für die Situation einer diffusen Start-Priori-Verteilung.

Wir wollen uns auf den Fall (ii) konzentrieren. De Jong (1989, Abschnitt 6) stellt einen (KFG)-Algorithmus für eine diffuse Start-Priori-Verteilung vor. Die Herleitung findet sich in De Jong (1991). Das Verfahren ist im Vergleich zum (KFG) mit informativer Start-Priori aus numerischer Sicht wesentlich aufwendiger: Es ist nicht mehr eine, sondern es sind drei Matrixrekursionen, die über den Zeitindex t ($t = 0, \dots, \bar{T}$) laufen.

Unser Ziel ist es, einen diffusen Algorithmus anzugeben, dessen Herleitung einfach und konstruktiv ist und für dessen Implementierung bestehende (KFG)-Software verwendet werden kann. Außerdem soll der numerische Aufwand dieses Verfahrens nicht größer als beim (KFG)-Algorithmus für eine informative Start-Priori sein. Für die Herleitung beschränken wir uns zunächst auf das lineare Zustandsraummodell mit dem zugehörigen Schätzverfahren (KFG). In den nächsten Teilabschnitten 5.2.1, 5.2.2 und 5.2.3 werden wir aus dem diffusen (KFG) den diffusen (LGKFG), (EKFG) und (IGKFG) ableiten.

Zur Vereinfachung nehmen wir an, daß im Übergangsmodell (3.2) für alle Komponenten von α_t (Trend, Saisonkomponenten, Kovariableneffekte) stets die gleiche Anzahl an Hilfsvariablen verwendet wurde, um die zugrunde liegenden strukturellen Zeitreihenmodelle in Zustandsraumform zu bringen. Beim RW1 beispielsweise benötigt man keine, beim RW2 eine Hilfsvariable. Verallgemeinerungen auf den Fall, daß zum Beispiel die Parameterübergänge für die Trendkomponente mit dem RW2 und für die Kovariableneffekte mit dem RW1 modelliert werden, sind ohne Einschränkung durch Partitionierung des Vektors α_t und getrennte Rekursionsformeln für jede Partition möglich. Außerdem sollen die Übergangsmatrizen F_t , $t = 1, \dots, \bar{T}$, stets regulär sein.

Falls der Parameterübergang in (3.2) mit einem RW1 festgelegt ist, so lässt sich aus (KFG) auf kanonische Weise ein diffuser (KFG)-Algorithmus ableiten:

Nach (KFG) gilt:

$$\begin{aligned} (V_{1|0}^{-1}) &= (F_1 Q_0 F_1' + Q_1)^{-1} \stackrel{\text{Matrix-Inversionslemma}}{=} Q_1^{-1} - Q_1^{-1} F_1' (F_1 Q_1^{-1} F_1' + Q_0^{-1})^{-1} F_1 Q_1^{-1} = \\ &= Q_1^{-1} - Q_1^{-1} = 0_{n,n}. \end{aligned}$$

Damit ist

$$V_{1|1} = (Z_1' \Sigma_1^{-1} Z_1)^{-1}. \quad (5.7)$$

Ferner gilt nach dem Korrekturschritt in (KFG) für $t = 1$:

$$a_{1|1} = a_{1|0} + V_{1|1} Z_1' \Sigma_1^{-1} (y_1 - Z_1 a_{1|0}) \stackrel{(5.7)}{=} a_{1|0} + V_{1|1} Z_1' \Sigma_1^{-1} y_1 - I_n a_{1|0} = V_{1|1} Z_1' \Sigma_1^{-1} y_1.$$

Damit erhalten wir insgesamt den diffusen (KFG) für den RW1 zu:

Diffuser (KFG) für RW1:

Initialisierung:

$$V_{1|1} = (Z_1' \Sigma_1^{-1} Z_1)^{-1},$$

$$a_{1|1} = V_{1|1} Z_1' \Sigma_1^{-1} y_1.$$

Für $t = 2, \dots, \bar{T}$:

Vorhersageschritt: Wie in (KFG).

Korrekturschritt: Wie in (KFG).

Für $t = \bar{T}, \dots, 2$:

Glättungsschritt: Wie in (KFG).

Mit $\alpha_{\bar{T}}^+ = (\alpha_1', \alpha_2', \dots, \alpha_{\bar{T}}')'$ löst der diffuse (KFG) das folgende quadratische Optimierungsproblem:

$$\begin{aligned}
\mathbf{a}^+ &:= (\mathbf{a}'_{1|\bar{T}}, \dots, \mathbf{a}'_{T|\bar{T}})' = \\
&= \arg \max_{\alpha_{\bar{T}}^+} \left\{ -\frac{1}{2} \sum_{t=1}^{\bar{T}} (\mathbf{y}_t - \mathbf{Z}_t \alpha_t)' \Sigma_t^{-1} (\mathbf{y}_t - \mathbf{Z}_t \alpha_t) - \frac{1}{2} \sum_{t=2}^{\bar{T}} (\alpha_t - \mathbf{F}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} (\alpha_t - \mathbf{F}_t \alpha_{t-1}) \right\},
\end{aligned} \tag{5.8}$$

denn das Optimierungsproblem in (4.7) ist im diffusen Fall äquivalent zu

$$\begin{aligned}
\mathbf{a} &\equiv (\mathbf{a}'_{0|\bar{T}}, \mathbf{a}'_{1|\bar{T}}, \dots, \mathbf{a}'_{T|\bar{T}})' \\
&\arg \max_{\alpha_{\bar{T}}^*} \left\{ -\frac{1}{2} \sum_{t=1}^{\bar{T}} (\mathbf{y}_t - \mathbf{Z}_t \alpha_t)' \Sigma_t^{-1} (\mathbf{y}_t - \mathbf{Z}_t \alpha_t) - \frac{1}{2} \sum_{t=1}^{\bar{T}} (\alpha_t - \mathbf{F}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} (\alpha_t - \mathbf{F}_t \alpha_{t-1}) \right\}.
\end{aligned} \tag{5.9}$$

Für eine optimale Lösung in (5.9) gilt jedoch insbesondere:

$$\mathbf{a}_{1|\bar{T}} = \mathbf{F}_1 \mathbf{a}_{0|\bar{T}}, \tag{5.10}$$

und damit ist (5.9) auch äquivalent zu (5.8).

Der diffuse (KFG) ist bei einem Parameterübergang, der mit dem RW2 modelliert ist, nicht anwendbar, weil $\mathbf{Z}'_1 \Sigma_1^{-1} \mathbf{Z}_1$ nicht invertierbar und somit $\mathbf{V}_{1|1}$ nicht ausrechenbar ist.

Ein anderes Schätzverfahren, bei dem der Einbezug einer diffusen Start-Priori-Verteilung problemlos möglich ist, liegt mit dem Informationsfilter vor. Er ist zum Beispiel in Anderson und Moore (1979, Abschnitt 6.3) angegeben. Wir stellen im folgenden die Rekursionsformeln des Informationsfilters mit diffuser Start-Priori explizit dar, weil wir sie später bei der Herleitung unseres Verfahrens noch benötigen. Die dabei auftretenden Größen $\mathbf{m}_{t|t-1}$ und $\mathbf{m}_{t|t}$ sind in bezug auf die im (KFG) auftretenden Größen $\mathbf{a}_{t|t-1}$ und $\mathbf{a}_{t|t}$ transformierte Vektoren, genauer

$$\mathbf{m}_{t|t-1} = \mathbf{V}_{t|t-1}^{-1} \mathbf{a}_{t|t-1} \text{ und } \mathbf{m}_{t|t} = \mathbf{V}_{t|t}^{-1} \mathbf{a}_{t|t}. \tag{5.11}$$

Informationsfilter:

Initialisierung: $V_{|0}^{-1} = V_{0|0}^{-1} = Q_0^{-1} = 0_{n,n}$, $m_{|0} = m_{0|0} = 0_n$,

$$V_{|1}^{-1} = Z_1' \Sigma_1^{-1} Z_1 \text{ , } m_{|1} = Z_1' \Sigma_1^{-1} y_1 .$$

Für $t = 2, \dots, \bar{T}$:

Vorhersageschritt: $A_t = (F_t^{-1})' V_{t-1|t-1}^{-1} F_t^{-1}$,

$$V_{t|t-1}^{-1} = \{I_n - A_t (A_t + Q_t^{-1})^{-1}\} A_t ,$$
$$m_{t|t-1} = \{I_n - A_t (A_t + Q_t^{-1})^{-1}\} (F_t^{-1})' m_{t-1|t-1} .$$

Korrekturschritt: $V_{t|t}^{-1} = V_{t|t-1}^{-1} + Z_t' \Sigma_t^{-1} Z_t$,

$$m_{t|t} = m_{t|t-1} + Z_t' \Sigma_t^{-1} y_t .$$

Wir sind jedoch nicht an den Werten $m_{t|t-1}$ und $m_{t|t}$, sondern an $a_{t|\bar{T}}$, $t = 1, \dots, \bar{T}$, interessiert. Die Idee ist es nun, den Informationsfilter, bei dem der Einbezug einer diffusen Start-Priori problemlos (auch zum Beispiel für den RW2) möglich ist, mit dem (KFG), der die Schätzungen $a_{t|\bar{T}}$, $t = 1, \dots, \bar{T}$, liefert, zu kombinieren. Dazu führen wir so lange Informationsfilterrekursionen, $t = 1, \dots, t_z$, aus, bis die Matrix $(V_{t_z+1|t_z+1}^{-1})^{-1}$ invertierbar ist. Für den RW1 ist $t_z = 0$. Setzt man die Modellspezifikationen für den RW2, die in (3.5) angegeben sind, in den Informationsfilter ein, so ergibt sich $t_z = 1$. Wir nehmen für unsere Übergangsmodelle in (3.2) an, daß t_z gleichzeitig die Anzahl der Hilfsvariablen ist, die benötigt werden, um das jeweilige strukturelle Zeitreihenmodell in Zustandsraumform zu bringen, was für den RW1, den RW2, den „random walk“ höherer Ordnung und für viele Saisonmodelle zutrifft. Für die optimale Lösung $a \equiv (a'_{0|\bar{T}}, a'_{1|\bar{T}}, \dots, a'_{\bar{T}|\bar{T}})'$ von (5.9) muß daher gelten:

$$\mathbf{a}_{t|\bar{T}} = \mathbf{F}_t \mathbf{a}_{t-1|\bar{T}}, \quad t = 1, \dots, t_z + 1. \quad (5.12)$$

Die Bedingung in (5.12) kann direkt in das lineare Zustandsraummodell (3.1), (3.2) eingebunden werden, indem man

$$\mathbf{Q}_t = \mathbf{0}_{n,n} \quad \text{für } t = 1, \dots, t_z + 1 \quad (5.13)$$

setzt. Man läßt also in (3.2) keine zufällige Schwankung zu und es ist $\text{pr}(\xi_t = 0) = 1, t = 1, \dots, t_z + 1$.

Aus numerischer Sicht bedeutet (5.12), daß das quadratische Optimierungsproblem in (5.9) äquivalent ist zu

$$\begin{aligned} \mathbf{a}^\times &:= (\mathbf{a}'_{t_z+1|\bar{T}}, \dots, \mathbf{a}'_{\bar{T}|\bar{T}})' \\ \arg \max_{\alpha^\times_{\bar{T}}} &\left\{ -\frac{1}{2} \sum_{t=1}^{\bar{T}} (\mathbf{y}_t - \mathbf{Z}_t \alpha_t)' \Sigma_t^{-1} (\mathbf{y}_t - \mathbf{Z}_t \alpha_t) - \frac{1}{2} \sum_{t=t_z+2}^{\bar{T}} (\alpha_t - \mathbf{F}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} (\alpha_t - \mathbf{F}_t \alpha_{t-1}) \right\}, \end{aligned} \quad (5.14)$$

mit $\alpha^\times_{\bar{T}} = (\alpha'_{t_z+1}, \alpha'_{t_z+2}, \dots, \alpha'_{\bar{T}})'$.

Im folgenden leiten wir die Rekursionsformeln für einen diffusen (KFG)-Algorithmus, der nicht nur den RW1, sondern auch allgemeine Übergangsmodelle der Form (3.2) zuläßt. Die Initialisierung erfolgt wie in oben beschriebenem Informationsfilter. Damit wird dem Vorliegen einer diffusen Start-Priori-Verteilung Rechnung getragen. Mit dem Matrix-Inversions-Lemma (Anderson und Moore, 1979, Abschnitt 6.3) erhält man für $\mathbf{V}_{t|t-1}^{-1}$, $t = 2, \dots, t_z + 1$ aus dem Informationsfilter

$$\mathbf{V}_{t|t-1}^{-1} = (\mathbf{A}_t^{-1} + \mathbf{Q}_t)^{-1}$$

und wegen (5.13)

$$\mathbf{V}_{t|t-1}^{-1} = \mathbf{A}_t. \quad (5.15)$$

Nach (5.11), dem Vorhersageschritt in (KFG) und der Definition von \mathbf{A}_t im Informationsfilter ist

$$\begin{aligned} \mathbf{m}_{t|t-1} &= \mathbf{V}_{t|t-1}^{-1} \mathbf{a}_{t|t-1} = \mathbf{V}_{t|t-1}^{-1} \mathbf{F}_t \mathbf{a}_{t-1|t-1} = \mathbf{A}_t \mathbf{F}_t \mathbf{a}_{t-1|t-1} = (\mathbf{F}_t^{-1})' \mathbf{V}_{t-1|t-1}^{-1} \mathbf{a}_{t-1|t-1} = \\ &= (\mathbf{F}_t^{-1})' \mathbf{m}_{t-1|t-1}, \quad t = 2, \dots, t_z + 1. \end{aligned} \quad (5.16)$$

Für den Glättungsschritt in (KFG) mit $t = 2, \dots, t_z + 1$ gilt

$$B_t = V_{t-1|t-1} F_t' V_{t|t-1}^{-1} \stackrel{(5.15)}{=} V_{t-1|t-1} F_t' (F_t^{-1})' V_{t-1|t-1}^{-1} F_t^{-1} = F_t^{-1} \quad (5.17a)$$

Damit erhalten wir für $t = 2, \dots, t_z + 1$:

$$a_{t-1|\bar{T}} = a_{t-1|t-1} + B_t (a_{t|\bar{T}} - a_{t|t-1}) = a_{t-1|t-1} + F_t^{-1} (a_{t|\bar{T}} - F_t a_{t-1|t-1}) = F_t^{-1} a_{t|\bar{T}}. \quad (5.17b)$$

Unsere Schätzung erfüllt also die Nebenbedingung $a_{t|\bar{T}} = F_t a_{t-1|\bar{T}}$, $t = 2, \dots, t_z + 1$, in (5.12).

Außerdem gilt für den Glättungsschritt in (KFG) mit $t = 2, \dots, t_z + 1$:

$$\begin{aligned} V_{t-1|\bar{T}} &= V_{t-1|t-1} + B_t (V_{t|\bar{T}} - V_{t|t-1}) B_t' = B_t V_{t|\bar{T}} B_t' + V_{t-1|t-1} - V_{t-1|t-1} F_t' V_{t|t-1}^{-1} F_t V_{t-1|t-1} \\ &\stackrel{(5.15)}{=} B_t V_{t|\bar{T}} B_t' + V_{t-1|t-1} - V_{t-1|t-1} \stackrel{(5.17a)}{=} F_t^{-1} V_{t|\bar{T}} (F_t^{-1})'. \end{aligned} \quad (5.18)$$

Damit lautet der diffuse Kalman Filter und Glätter:

Diffuser (KFG):

Initialisierung: $V_{1|1}^{-1} = Z_1' \Sigma_1^{-1} Z_1$, $m_{1|1} = Z_1' \Sigma_1^{-1} y_1$.

Für $t = 2, \dots, t_z + 1$:

$$V_{t|t-1}^{-1} = (F_t^{-1})' V_{t-1|t-1}^{-1} F_t^{-1},$$

$$m_{t|t-1} = (F_t^{-1})' m_{t-1|t-1},$$

$$V_{t|t}^{-1} = V_{t|t-1}^{-1} + Z_t' \Sigma_t^{-1} Z_t,$$

$$m_{t|t} = m_{t|t-1} + Z_t' \Sigma_t^{-1} y_t.$$

$$V_{t_z+1|t_z+1} = (V_{t_z+1|t_z+1}^{-1})^{-1}, \quad a_{t_z+1|t_z+1} = V_{t_z+1|t_z+1} m_{t_z+1|t_z+1}.$$

Für $t = t_z + 2, \dots, \bar{T}$:

Vorhersageschritt: Wie in (KFG).

Korrekturschritt: Wie in (KFG).

Für $t = \bar{T}, \dots, t_z + 2$:

Glättungsschritt: Wie in (KFG).

Für $t = t_z + 1, \dots, 2$:

$$\mathbf{a}_{t-1|\bar{T}} = \mathbf{F}_t^{-1} \mathbf{a}_{t|\bar{T}},$$

$$\mathbf{V}_{t-1|\bar{T}} = \mathbf{F}_t^{-1} \mathbf{V}_{t|\bar{T}} (\mathbf{F}_t^{-1})'.$$

Falls die Matrizen \mathbf{Z}_t und \mathbf{F}_t , $t = 1, \dots, \bar{T}$, regulär sind, dann ist der diffuse (KFG) äquivalent mit dem in De Jong (1991) angegebenen Verfahren. Das Ergebnis \mathbf{a}^\times , das wir mit dem diffusen (KFG) erhalten, löst das quadratische Optimierungsproblem in (5.14). Wie im Falle einer informativen Start-Priori-Verteilung leiten wir nun aus dem diffusen (KFG) den diffusen (LGKFG) ab.

5.2.1 Diffuser linear gewichteter Kalman Filter und Glätter (D-LGKFG)

Mit der gleichen Begründung wie in Abschnitt 5.1 ersetzen wir die Beobachtungen \mathbf{y}_t durch die „working observation“ $\tilde{\mathbf{y}}_t(\mathbf{a}_t^{(i)})$ aus (4.25) und die inverse Fehler-Kovarianzmatrix Σ_t^{-1} aus dem Beobachtungsmodell (3.1) durch die Gewichtsmatrix $\mathbf{W}_t(\mathbf{a}_t^{(i)})$ für $t = 1, \dots, \bar{T}$ unter Beachtung der Definition von $\tilde{\mathbf{y}}(\mathbf{a}^{(i)})$ in (4.25) und von $\mathbf{W}(\mathbf{a}^{(i)})$ in (4.19). Dabei geht $\mathbf{Z}_t' \Sigma_t^{-1} \mathbf{y}_t$ über in

$$\mathbf{Z}_t' \mathbf{D}_t(\mathbf{a}_t^{(i)}) \Sigma_t^{-1}(\mathbf{a}_t^{(i)}) \mathbf{D}_t'(\mathbf{a}_t^{(i)}) [\{\mathbf{D}_t^{-1}(\mathbf{a}_t^{(i)})\}' \{\mathbf{y}_t - \mu_t(\mathbf{a}_t^{(i)})\} + \mathbf{Z}_t \mathbf{a}_t^{(i)}] = \mathbf{s}_t(\mathbf{a}_t^{(i)}) + \mathbf{S}_t(\mathbf{a}_t^{(i)}) \cdot \mathbf{a}_t^{(i)}.$$

Insgesamt erhalten wir die folgenden Rekursionsformeln für den diffusen linear gewichteten Kalman Filter und Glätter:

(D-LGKFG):

Initialisierung: $V_{1|1}^{-1} = S_1(a_1^{(i)})$, $m_{1|1} = s_1(a_1^{(i)}) + V_{1|1}^{-1}a_1^{(i)}$.

Für $t = 2, \dots, t_z + 1$:

$$V_{t|t-1}^{-1} = (F_t^{-1})' V_{t-1|t-1}^{-1} F_t^{-1} ,$$

$$m_{t|t-1} = (F_t^{-1})' m_{t-1|t-1} ,$$

$$V_{t|t}^{-1} = V_{t|t-1}^{-1} + S_t(a_t^{(i)}) ,$$

$$m_{t|t} = m_{t|t-1} + s_t(a_t^{(i)}) + S_t(a_t^{(i)}) \cdot a_t^{(i)} .$$

$$V_{t_z+1|t_z+1} = (V_{t_z+1|t_z+1}^{-1})^{-1} , a_{t_z+1|t_z+1} = V_{t_z+1|t_z+1} m_{t_z+1|t_z+1} .$$

Für $t = t_z + 2, \dots, \bar{T}$:

Vorhersageschritt: Wie in (LGKFG).

Korrekturschritt: Wie in (LGKFG).

Für $t = \bar{T}, \dots, t_z + 2$:

Glättungsschritt: Wie in (LGKFG).

Für $t = t_z + 1, \dots, 2$:

$$a_{t-1|\bar{T}}^{(i+1)} = F_t^{-1} a_{t|\bar{T}}^{(i+1)} ,$$

$$V_{t-1|\bar{T}} = F_t^{-1} V_{t|\bar{T}} (F_t^{-1})' .$$

Als Lösung von (D-LGKFG) erhalten wir $(a^{(i+1)})^+ \in \mathbb{R}^{m-n}$, wobei $a^{(i+1)} = ((a_0^{(i+1)})', ((a^{(i+1)})^+)')'$ für den Fall einer diffusen Start-Priori-Verteilung. Wir haben also einen Fisher-Scoring-Schritt von $(a^{(i)})^+$ nach $(a^{(i+1)})^+$ bei diffuser Start-Priori-Verteilung ausgeführt. Aus dem diffusen (LGKFG) leiten wir im nächsten Teilabschnitt die diffuse Version des (EKFG) ab.

5.2.2 Diffuser verallgemeinerter und erweiterter Kalman Filter und Glätter (D-EKFG)

Mit

$$(\mathbf{a}^{(i)})^+ = (\mathbf{a}'_{1|0}, \mathbf{a}'_{2|1}, \dots, \mathbf{a}'_{t|t-1}, \dots, \mathbf{a}'_{\bar{T}|\bar{T}-1})', \quad (5.19)$$

in (D-LGKFG) eingesetzt ergibt sich analog zum (EKFG) in Teilabschnitt 5.1.2 der diffuse verallgemeinerte und erweiterte Kalman Filter und Glätter:

(D-EKFG):

Initialisierung: Wähle $\mathbf{a}_{1|0}$ beliebig $\in \mathbb{R}^{m-n}$ (beispielsweise $\mathbf{a}_{1|0} = \mathbf{0}_n$).

$$\mathbf{V}_{1|1}^{-1} = \mathbf{S}_1(\mathbf{a}_{1|0}), \quad \mathbf{m}_{1|1} = \mathbf{s}_1(\mathbf{a}_{1|0}) + \mathbf{V}_{1|1}^{-1} \mathbf{a}_{1|0}.$$

Für $t = 2, \dots, t_z + 1$:

$$\text{Setze } \mathbf{a}_{t|t-1} = \mathbf{F}_t \mathbf{a}_{t-1|t-2}.$$

(Falls $\mathbf{a}_{1|0} = \mathbf{0}_n$ gewählt wurde, dann ist $\mathbf{a}_{t|t-1} = \mathbf{0}_n$.)

$$\mathbf{V}_{t|t-1}^{-1} = (\mathbf{F}_t^{-1})' \mathbf{V}_{t-1|t-1}^{-1} \mathbf{F}_t^{-1},$$

$$\mathbf{m}_{t|t-1} = (\mathbf{F}_t^{-1})' \mathbf{m}_{t-1|t-1},$$

$$\mathbf{V}_{t|t}^{-1} = \mathbf{V}_{t|t-1}^{-1} + \mathbf{S}_t(\mathbf{a}_{t|t-1}),$$

$$\mathbf{m}_{t|t} = \mathbf{m}_{t|t-1} + \mathbf{s}_t(\mathbf{a}_{t|t-1}) + \mathbf{S}_t(\mathbf{a}_{t|t-1}) \cdot \mathbf{a}_{t|t-1}.$$

$$\mathbf{V}_{t_z+1|t_z+1} = (\mathbf{V}_{t_z+1|t_z+1}^{-1})^{-1}, \quad \mathbf{a}_{t_z+1|t_z+1} = \mathbf{V}_{t_z+1|t_z+1} \mathbf{m}_{t_z+1|t_z+1}.$$

Für $t = t_z + 2, \dots, \bar{T}$:

Vorhersageschritt: Wie in (EKFG).

Korrekturschritt: Wie in (EKFG).

Für $t = \bar{T}, \dots, t_z + 2$:

Glättungsschritt: Wie in (EKFG).

Für $t = t_z + 1, \dots, 2$:

$$\mathbf{a}_{t-1|\bar{T}}^{(1)} = \mathbf{F}_t^{-1} \mathbf{a}_{t|\bar{T}}^{(1)},$$

$$\mathbf{V}_{t-1|\bar{T}} = \mathbf{F}_t^{-1} \mathbf{V}_{t|\bar{T}} (\mathbf{F}_t^{-1})'.$$

Als Ergebnis erhalten wir $(\mathbf{a}^{(1)})^+ = ((\mathbf{a}_1^{(1)})', \dots, (\mathbf{a}_{\bar{T}}^{(1)})')' \in \mathbb{R}^{m-n}$ bei Vorliegen einer diffusen Start-Priori. Wir werden den (D-EKFG) als Startiteration für unser diffuses Fisher-Scoring-Verfahren im nachfolgenden Teilabschnitt verwenden.

5.2.3 Diffuser iterativ gewichteter Kalman Filter und Glätter (D-IGKFG)

Um analog zu der Vorgehensweise in Teilabschnitt 5.1.3 zu einem Fisher-Scoring-Verfahren zu kommen, kombinieren wir den (D-EKGF) mit (D-LGKFG)-Iterationen und erhalten den diffusen iterativ gewichteten Kalman Filter und Glätter:

Diffuser iterativ gewichteter Kalman Filter und Glätter (D-IGKFG):

Initialisierung: Berechne $(\mathbf{a}^{(1)})^+ = ((\mathbf{a}_1^{(1)})', \dots, (\mathbf{a}_{\bar{T}}^{(1)})')' \in \mathbb{R}^{m-n}$ mit (D-EKFG).

Setze den Iterationsindex $i = 1$.

Schritt 1: Mit dem Eingangsvektor $(\mathbf{a}^{(i)})^+ \in \mathbb{R}^{m-n}$ berechne $(\mathbf{a}^{(i+1)})^+ \in \mathbb{R}^{m-n}$ unter Verwendung von (D-LGKFG).

Schritt 2: Wenn ein Abbruchkriterium erfüllt ist: STOP.

Anderenfalls erhöhe den Iterationsindex i um 1 und gehe zu Schritt 1.

Als Ergebnis von (D-IGKFG) bekommt man $\mathbf{a}^+ := (\mathbf{a}'_{1|\bar{T}}, \dots, \mathbf{a}'_{\bar{T}|\bar{T}})'$ mit

$$\mathbf{a}^+ = \arg \max_{\alpha_{\bar{T}}^+} \left\{ -\frac{1}{2} \sum_{t=1}^{\bar{T}} \ln p(y_t | \alpha_t, y_{t-1}^*, \mathbf{x}_t^*) - \frac{1}{2} \sum_{t=2}^{\bar{T}} (\alpha_t - F_t \alpha_{t-1})' Q_t^{-1} (\alpha_t - F_t \alpha_{t-1}) \right\}, \quad (5.20)$$

wobei die Lösung des nichtlinearen Optimierungsproblems in (5.20) für den diffusen Fall äquivalent ist mit dem Lösen von (4.5). Wir sprechen vom diffusen Fall, wenn die Verteilung von α_0 in (3.2) beziehungsweise (3.7) als nichtinformativ angenommen wird. Wegen (5.12) ist (5.20) auch äquivalent zu dem in

$$\mathbf{a}^\times = \arg \max_{\alpha_{\bar{T}}^\times} \left\{ -\frac{1}{2} \sum_{t=1}^{\bar{T}} \ln p(y_t | \alpha_t, y_{t-1}^*, \mathbf{x}_t^*) - \frac{1}{2} \sum_{t=t_z+2}^{\bar{T}} (\alpha_t - F_t \alpha_{t-1})' Q_t^{-1} (\alpha_t - F_t \alpha_{t-1}) \right\} \quad (5.21)$$

enthaltenen Optimierungsproblem. (5.21) kann (mit dem RW2 im Übergangsmodell) als zeitdiskretes Analogon zu nonparametrischen Glättungskriterien (Spline-Glättung) gesehen werden. In beiden Fällen kommt der „künstliche“ Startparameter α_0 nicht vor.

Der (D-IGKFG)-Algorithmus ist ein komplettes Fisher-Scoring-Verfahren bei diffuser Start-Priori-Verteilung. Die Fehler-Kovarianzmatrizen Q_t , $t = 1, \dots, \bar{T}$, werden an keiner Stelle explizit invertiert. Somit können auch Übergangsmodelle wie der RW2 verwendet werden, bei denen die Matrizen Q_t singulär sind. Auch zeitkonstante Parameter $\alpha = \alpha_0 = \alpha_1 = \dots = \alpha_{\bar{T}}$ mit diffuser Verteilung von α_0 lassen sich mit (D-IGKFG) und der Wahl $Q_1 = Q_2 = \dots = Q_{\bar{T}} = \mathbf{0}_{n,n}$ schätzen.

Aus numerischer Sicht ist der (D-IGKFG) ebenso effizient wie der (IGKFG): Die erwartete Fisher-Informationsmatrix in (5.21) wird nicht invertiert und ihre blocktridiagonale Gestalt im Lösungsverfahren explizit berücksichtigt und ausgenützt.

Wie der verbleibende Hyperparameter Q_t zu schätzen ist, zeigt der nächste Teilabschnitt.

5.2.4 EM-Typ Algorithmus für diffuse Start-Priori

Die Start-Hyperparameter \mathbf{a}_0 und Q_0 treten im diffusen Fall nicht mehr auf und kommen im (D-IGKFG) nicht vor. Zu schätzen bleibt, abgesehen von den Zuständen α_t , der Hyperparameter Q , wobei wir analog zu Teilabschnitt 5.1.4 annehmen, daß $Q = Q_{t_z+2} = Q_{t_z+3} = \dots = Q_{\bar{T}}$. Für Q_t , $t = 1, \dots, t_z + 1$, gilt ja nach (5.13):

$$\mathbf{Q}_t = \mathbf{0}_{n,n} \text{ für } t = 1, \dots, t_z + 1.$$

Die gemeinsame Loglikelihood der Daten ist, abgesehen von Konstanten, die \mathbf{Q} nicht enthalten, gegeben durch

$$\begin{aligned} l(y_1, \dots, y_{\bar{T}}, \alpha_{t_z+1}, \dots, \alpha_{\bar{T}}; \mathbf{Q}) = & -\frac{\bar{T} - t_z - 1}{2} \log(\det \mathbf{Q}) - \\ & -\frac{1}{2} \sum_{t=t_z+2}^{\bar{T}} (\alpha_t - \mathbf{F}_t \alpha_{t-1})' \mathbf{Q}_t^{-1} (\alpha_t - \mathbf{F}_t \alpha_{t-1}). \end{aligned} \quad (5.22)$$

Erwartungswertbildung $E(\mathbf{Q}) := \mathbb{E}l(y_1, \dots, y_{\bar{T}}, \alpha_{t_z+1}, \dots, \alpha_{\bar{T}}; \mathbf{Q}) | y_1, \dots, y_{\bar{T}}; \mathbf{Q}^{(p)}\}$ unter Beachtung von $\text{cov}(\alpha_{t-1} \alpha_t' | y_1, \dots, y_{\bar{T}}; \mathbf{Q}^{(p)}) = \mathbf{B}_t \mathbf{V}_{t|\bar{T}}$ nach De Jong und Mackinnon (1988) mit \mathbf{B}_t und $\mathbf{V}_{t|\bar{T}}$ wie in (D-IGKFG) und Maximierung von $E(\mathbf{Q})$ nach \mathbf{Q} ergibt $\mathbf{Q}^{(p+1)}$ zu

$$\begin{aligned} \mathbf{Q}^{(p+1)} = & \frac{1}{\bar{T} - t_z - 1} \sum_{t=t_z+2}^{\bar{T}} \left\{ \left(\mathbf{a}_{t|\bar{T}} - \mathbf{F}_t \mathbf{a}_{t-1|\bar{T}} \right) \left(\mathbf{a}_{t|\bar{T}} - \mathbf{F}_t \mathbf{a}_{t-1|\bar{T}} \right)' + \mathbf{V}_{t|\bar{T}} \right. \\ & \left. - \mathbf{F}_t \mathbf{B}_t \mathbf{V}_{t|\bar{T}} - \mathbf{V}_{t|\bar{T}}' \mathbf{B}_t' \mathbf{F}_t' + \mathbf{F}_t \mathbf{V}_{t-1|\bar{T}} \mathbf{F}_t' \right\}, \end{aligned} \quad (5.23)$$

mit \mathbf{B}_t wie im Glättungsschritt von (D-IGKFG). Für die Teilsumme

$$\bar{\mathbf{S}}(t_z + 1) := \sum_{t=1}^{t_z+1} \left\{ \left(\mathbf{a}_{t|\bar{T}} - \mathbf{F}_t \mathbf{a}_{t-1|\bar{T}} \right) \left(\mathbf{a}_{t|\bar{T}} - \mathbf{F}_t \mathbf{a}_{t-1|\bar{T}} \right)' + \mathbf{V}_{t|\bar{T}} - \mathbf{F}_t \mathbf{B}_t \mathbf{V}_{t|\bar{T}} - \mathbf{V}_{t|\bar{T}}' \mathbf{B}_t' \mathbf{F}_t' + \mathbf{F}_t \mathbf{V}_{t-1|\bar{T}} \mathbf{F}_t' \right\}$$

erhalten wir mit (5.12)

$$\bar{\mathbf{S}}(t_z + 1) = \sum_{t=1}^{t_z+1} \left\{ \mathbf{V}_{t|\bar{T}} - \mathbf{F}_t \mathbf{B}_t \mathbf{V}_{t|\bar{T}} - \mathbf{V}_{t|\bar{T}}' \mathbf{B}_t' \mathbf{F}_t' + \mathbf{F}_t \mathbf{V}_{t-1|\bar{T}} \mathbf{F}_t' \right\} \stackrel{(5.17a)}{=} \mathbf{0}_{n,n} \quad (5.24)$$

Das Ergebnis in (5.24) ist konsistent mit (5.13). Insgesamt erhalten wir zur simultanen Schätzung von \mathbf{Q} und den Zuständen α_t den

EM-Typ Algorithmus für diffuse Start-Priori:

Schritt 1: Wähle einen Startwert $Q^{(0)}$ für Q und setze den Iterationsindex $p = 0$.

Schritt 2: Berechne $a_{t|\bar{T}}, V_{t|\bar{T}}$ ($t = 1, \dots, \bar{T}$) mit (D-EKFG) oder (D-IGKFG), unter Verwendung von $Q^{(p)}$.

Schritt 3: EM-Typ Schritt: Berechne $Q^{(p+1)}$ mit den zuvor in Schritt 2 bestimmten Größen $a_{t|\bar{T}}, V_{t|\bar{T}}$ ($t = 1, \dots, \bar{T}$) durch

$$Q^{(p+1)} = \frac{1}{\bar{T} - t_z - 1} \sum_{t=t_z+2}^{\bar{T}} \left\{ \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right) \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right)' + V_{t|\bar{T}} - F_t B_t V_{t|\bar{T}} - V_{t|\bar{T}}' B_t' F_t' + F_t V_{t-1|\bar{T}} F_t' \right\},$$

mit B_t wie im Glättungsschritt von (D-IGKFG) oder (D-EKFG).

Schritt 4: Wenn ein Abbruchkriterium erreicht wurde: STOP.

Sonst erhöhe den Iterationsindex p um 1 und gehe zu Schritt 2.

5.2.5 EM-Typ Algorithmus für unterschiedliche Zeitintervalllängen

Im folgenden wollen wir den EM-Typ Algorithmus so modifizieren, daß er für das piecewise exponential Modell aus Abschnitt 3.2 bei diffuser Start-Priori-Verteilung geeignet ist. Dazu setzen wir voraus, daß $Q = Q_1 = Q_2 = \dots = Q_{\bar{T}}$ sein soll. Den einzigen Hyperparameter, den es also zu schätzen gilt, ist Q . Die gemeinsame Loglikelihood der Daten für das „diffuse“ piecewise exponential Modell ist, abgesehen von Konstanten, die Q nicht enthalten:

$$l(y_1, \dots, y_{\bar{T}}, \alpha_{t_z+1}, \dots, \alpha_{\bar{T}}; Q) = -\frac{1}{2} \sum_{t=t_z+2}^{\bar{T}} \log[\det\{w_t \cdot Q\}]$$

$$-\frac{1}{2} \sum_{t=t_z+2}^{\bar{T}} (\alpha_t - F_t \alpha_{t-1})' (c_t \cdot Q)^{-1} (\alpha_t - F_t \alpha_{t-1}). \quad (5.25)$$

w_t ist dabei die Länge des t -ten Zeitintervalls, $c_t = w_t$, falls im piecewise exponential Modell $w_t \cdot Q_t$ für die Kovarianzmatrix der Fehlerterme ξ_t , und $c_t = 1/w_t$, wenn $(1/w_t) \cdot Q_t$ für die Kovarianzmatrix der Fehlerterme ξ_t verwendet wurde. Nach algebraischen Umformungen erhält man

$$l(y_1, \dots, y_{\bar{T}}, \alpha_{t_z+1}, \dots, \alpha_{\bar{T}}; Q) = -\frac{1}{2} \sum_{t=t_z+2}^{\bar{T}} \log(n \cdot w_t) - \frac{\bar{T} - t_z - 1}{2} \log(\det Q) -$$

$$-\frac{1}{2} \sum_{t=t_z+2}^{\bar{T}} \frac{1}{c_t} (\alpha_t - F_t \alpha_{t-1})' Q^{-1} (\alpha_t - F_t \alpha_{t-1}). \quad (5.26)$$

Erwartungswertbildung und Maximierung wie im EM-Typ Algorithmus für diffuse Start-Priori ergibt unter Berücksichtigung der Linearität von Erwartungswert und Ableitung

$$Q^{(p+1)} = \frac{1}{\bar{T} - t_z - 1} \cdot \sum_{t=t_z+2}^{\bar{T}} \frac{1}{c_t} \left\{ \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right) \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right)' + V_{t|\bar{T}} \right.$$

$$\left. - F_t B_t V_{t|\bar{T}} - V_{t|\bar{T}}' B_t' F_t' + F_t V_{t-1|\bar{T}} F_t' \right\}. \quad (5.27)$$

Damit liegt der EM-Typ Algorithmus für unterschiedliche Zeitintervalllängen bei diffuser Start-Priori-Verteilung vor:

EM-Typ Algorithmus für unterschiedliche Zeitintervalllängen:

Schritt 1: Wähle einen Startwert $Q^{(0)}$ für Q und setze den Iterationsindex $p = 0$.

Schritt 2: Berechne $a_{t|\bar{T}}, V_{t|\bar{T}}$ ($t = 1, \dots, \bar{T}$) mit (D-EKFG) oder (D-IGKFG), unter Verwendung von $Q^{(p)}$. Beachte dabei die entsprechenden

Modellspezifikationen für das piecewise exponential Modell, insbesondere die Berechnung der Score-Funktion $s_t(\alpha_t)$ nach (4.43) und der erwarteten Informationsmatrix $S_t(\alpha_t)$ nach (4.44).

Schritt 3: EM-Typ Schritt: Berechne $Q^{(p+1)}$ mit den zuvor in Schritt 2 bestimmten Größen $a_{t|\bar{T}}, V_{t|\bar{T}}$ ($t = 1, \dots, \bar{T}$) durch

$$Q^{(p+1)} = \frac{1}{\bar{T} - t_z - 1} \cdot \sum_{t=t_z+2}^{\bar{T}} \frac{1}{c_t} \left\{ \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right) \left(a_{t|\bar{T}} - F_t a_{t-1|\bar{T}} \right)' + V_{t|\bar{T}} - F_t B_t V_{t|\bar{T}} - V_{t|\bar{T}}' B_t' F_t' + F_t V_{t-1|\bar{T}} F_t' \right\},$$

mit B_t wie im Glättungsschritt von (D-IGKFG) oder (D-EKFG).

Schritt 4: Wenn ein Abbruchkriterium erreicht wurde: STOP.

Sonst erhöhe den Iterationsindex p um 1 und gehe zu Schritt 2.

5.3 Numerische Aspekte

Die Inversion einer Matrix mit der Dimension $n \times n$ ist nicht nur eine aufwendige Operation (Größenordnung n^3 Operationen), sondern oft auch erheblich mit numerischen Fehlern behaftet. Bei der Berechnung von $V_{t|t}$ und $a_{t|t}$ im Korrekturschritt von (D-LGKFG), (D-EKFG), (LGKFG), und (EKFG) müssen drei Matrizen invertiert werden, wobei sich eine Matrixinvertierung stets durch äquivalente Umformung und eine zweite in besonderen Fällen vermeiden läßt.

In Anhang B beweisen wir durch Äquivalenzumformungen, daß die „Filterkovarianzmatrizen“ $V_{t|t}$ im Korrekturschritt auch berechnet werden können nach der Formel

$$V_{t|t} = (I_n + V_{t|t-1} S_t(\alpha_t))^{-1} V_{t|t-1} . \quad (5.28)$$

α_t in (5.28) ist je nach Algorithmus mit den dort angegebenen Vektoren beziehungsweise Zwischenwerten zu belegen. Eine weitere Möglichkeit wäre die Berechnung nach der

Kalman-Gain-Form. Insgesamt muß der Benutzer selbst entscheiden, welche Variante für sein vorliegendes Problem die numerisch günstigere ist.

Im Falle des multinomialen Zustandsraummodells muß die Inversenoperation auf $\Sigma_t(\alpha_t)$, der Kovarianzmatrix aus dem Beobachtungsmodell, für die Berechnung der Score-Funktion $s_t(\alpha_t)$ nach (4.15) oder (4.43) und die Bestimmung der erwarteten Informationsmatrix $S_t(\alpha_t)$ nach (4.17) oder (4.44) nicht angewendet werden. Mit (4.45) liegt eine Formel für $\Sigma_t^{-1}(\alpha_t)$ vor.

Bei der Verwendung der natürlichen Response-Funktion $h(\eta)$ im Beobachtungsmodell, wobei η der lineare Prädiktor ist, läßt sich ebenfalls die Invertierung von $\Sigma_t(\alpha_t)$ vermeiden. Es gilt dann:

$$\frac{\partial h(\eta)}{\partial \eta} = V(\eta), \quad (5.29)$$

und $V(\eta)$ ist die Varianzfunktion der zu $h(\eta)$ gehörigen Exponentialfamilienverteilung. Für die Varianzfunktion wiederum gilt: $\Sigma_t(\alpha_t) = C(\theta, \omega_{\text{W}}) \cdot V(\eta)$, wobei $C(\theta, \omega_{\text{W}}) \in \mathbb{R}$ lediglich vom natürlichen Parameter θ und vom Gewichtsparameter ω_{W} der Exponentialfamilienverteilung abhängt. Mit (5.29) erhalten wir insgesamt für die Score-Funktion in (4.15) oder (4.43) und die erwartete Fisher-Informationsmatrix in (4.17) oder (4.44):

$$\begin{aligned} s_t(\alpha_t) &= C(\theta, \omega_{\text{W}}) \cdot Z_t' \{y_t - \mu_t(\alpha_t)\} \\ S_t(\alpha_t) &= C(\theta, \omega_{\text{W}}) \cdot Z_t' D_t(\alpha_t) Z_t. \end{aligned} \quad (5.30)$$

Damit ist in diesem Falle die Berechnung von $\Sigma_t(\alpha_t)$ bei Verwendung der Algorithmen (D-LGKFG), (D-EKFG), (LGKFG), oder (EKFG) nicht notwendig.

Im Falle der natürlichen Response-Funktion $h(\eta)$ ergibt sich ein weiterer, algorithmischer Vorteil: Die erwartete Informationsmatrix ist hier identisch mit der beobachteten Informationsmatrix (siehe beispielsweise Fahrmeir und Tutz, 1994, Anhang A1). Damit ist das Fisher-Scoring-Verfahren identisch mit dem Newton-Verfahren aus der nichtlinearen Optimierung, für das man lokal quadratische Konvergenzgeschwindigkeit beweisen kann (vergleiche etwa Gill, Murray und Wright, 1981). Das bedeutet, wenn man die natürliche Responsefunktion im Beobachtungsmodell verwendet und eine „gute Näherung“ $\mathbf{a}^{(i)}$ für die Lösung, den Posteriori-Modus-Schätzer \mathbf{a} , besitzt, dann wird der (IGKFG)- und auch der

(D-IGKFG)-Algorithmus sehr schnell, mit der Größenordnung von zwei gültigen Stellen pro Iteration, gegen die Lösung a konvergieren.

6 Datenanalyse

Für die graphische Darstellung der Schätzergebnisse geben wir im folgenden stets punktweise Konfidenzbänder an. Liegt beispielsweise das untere Konfidenzband eines geschätzten Effektes über dem Nullniveau, so bezeichnen wir den Effekt als signifikant positiv. Soll die i -te Komponente $\mathbf{a}_{t\bar{T}}^i$ der Schätzung $\mathbf{a}_{t\bar{T}} = (\mathbf{a}_{t\bar{T}}^1, \mathbf{a}_{t\bar{T}}^2, \dots, \mathbf{a}_{t\bar{T}}^n)'$ dargestellt werden, so geben wir das zugehörige Konfidenzintervall

$$\mathbf{a}_{t\bar{T}}^i \pm \sqrt{\mathbf{V}_{t\bar{T}}^i} \quad (6.1)$$

an. $\mathbf{V}_{t\bar{T}}^i$ ist dabei das i -te Element in der Hauptdiagonalen von $\mathbf{V}_{t\bar{T}}$. Zur Angabe der Konfidenzintervalle bei den geschätzten Hazardraten $\hat{\lambda}_t = \mathbf{h}(\hat{\eta}_t)$ verwenden wir die δ -Methode (Fahrmeir und Hamerle, 1984, Satz 3.13, Seite 33) :

$$\text{var}(\hat{\lambda}_t) = \left(\frac{\partial \mathbf{h}(\hat{\eta}_t)}{\partial \eta_t} \right)' \mathbf{Z}_t \mathbf{V}_{t\bar{T}} \mathbf{Z}_t' \left(\frac{\partial \mathbf{h}(\hat{\eta}_t)}{\partial \eta_t} \right). \quad (6.2)$$

Das Konfidenzintervall für das Zeitintervall t ist somit gegeben durch

$$\hat{\lambda}_t^i \pm \sqrt{\{\text{var}(\hat{\lambda}_t)\}^i}, \quad (6.4)$$

wobei $\hat{\lambda}_t^i$ die i -te Komponente von $\hat{\lambda}_t$ und $\{\text{var}(\hat{\lambda}_t)\}^i$ das i -te Hauptdiagonalelement von $\text{var}(\hat{\lambda}_t)$ ist.

Als Schätzverfahren wurde stets, wenn nicht anders angegeben, der EM-Typ Algorithmus für eine diffuse Start-Priori aus Teilabschnitt 5.2.4 verwendet. Als Unterprogramm haben wir dabei den (D-IGKFG) benutzt.

6.1 Magenkrebsdaten

Im folgenden analysieren wir die Magenkrebsdaten, die in Teilabschnitt 1.1.1 bereits vorgestellt wurden. Zunächst geschieht dies im Rahmen eines dynamischen binomialen Logit-Modells. Um unterschiedliche Zeitintervalllängen zu berücksichtigen und die Wahl verschiedener Zeitgrids zu untersuchen, verwenden wir das piecewise exponential Modell.

6.1.1 Dynamisches binomiales Logit-Modell

Im Rahmen des dynamischen binomialen Logit-Modells haben wir die Daten auf drei Arten untersucht:

- 1.) Subgruppenanalyse,
- 2.) Gesamtanalyse, bei der alle Untersuchungseinheiten einbezogen sind, mit dem RW1, und schließlich
- 3.) Gesamtanalyse mit dem RW2.

Für die Analysen in diesem Teilabschnitt wurden die Tagesdaten in Monatsdaten gruppiert, das heißt zu Gruppen von jeweils 30 Tagen zusammengefaßt. Damit liegen pro Zeitintervall mehr Beobachtungen und somit auch mehr Information vor.

Bei der Subgruppenanalyse wurden nur jeweils die Daten der ersten Behandlungsgruppe (Chemo- und Strahlentherapie) beziehungsweise der zweiten (nur Chemotherapie) benutzt. Dabei haben wir als Beobachtungsmodell ein dynamisches binomiales Logit-Modell nach (3.14) zugrunde gelegt:

$$\lambda(t) = \frac{\exp(\alpha_t)}{1 + \exp(\alpha_t)}, \quad t = 1, \dots, \bar{T} = 56. \quad (6.5)$$

Als Übergangsmodell für den Baseline-Parameter α_t mit $n = \dim(\alpha_t) = 1$ haben wir den RW1 aus (3.4) verwendet, das heißt

$$\alpha_t = \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(0, Q), \quad t = 1, \dots, \bar{T} = 56.$$

Verteilung von α_0 : diffus. (6.6)

Tabelle 6.1 gibt über den algorithmischen Ablauf der EM-Typ-Schätzung für die erste Therapiegruppe Auskunft. $Q^{(0)}$ ist dabei wie beim EM-Typ-Algorithmus der Startwert für die Hyperparameterschätzung von Q und \hat{Q} das Ergebnis der Schätzung. FS-Iter. bezeichnet die Anzahl der Fisher-Scoring-Iterationen, FS_acc. die Abbruchschranke für das Fisher-Scoring-Verfahren, und EM-Iter. beziehungsweise EM_acc. entsprechend für den EM-Typ-Algorithmus.

Magenkrebsdaten: Subgruppenanalyse: 1.Therapiegruppe					
$Q^{(0)}$	\hat{Q}	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.014	0.014169	6	$1 \cdot 10^{-10}$	1	$2 \cdot 10^{-4}$

Tabelle 6.1: Hyperparameterschätzung und Programmablauf von PME.ST.

Der „gute“ Startwert für den Hyperparameter Q war bereits aus zahlreichen Programm-
läufen mit informativer Start-Priori-Verteilung bekannt, so daß der EM-Typ Algorithmus
nur noch eine EM-Iteration durchführte.

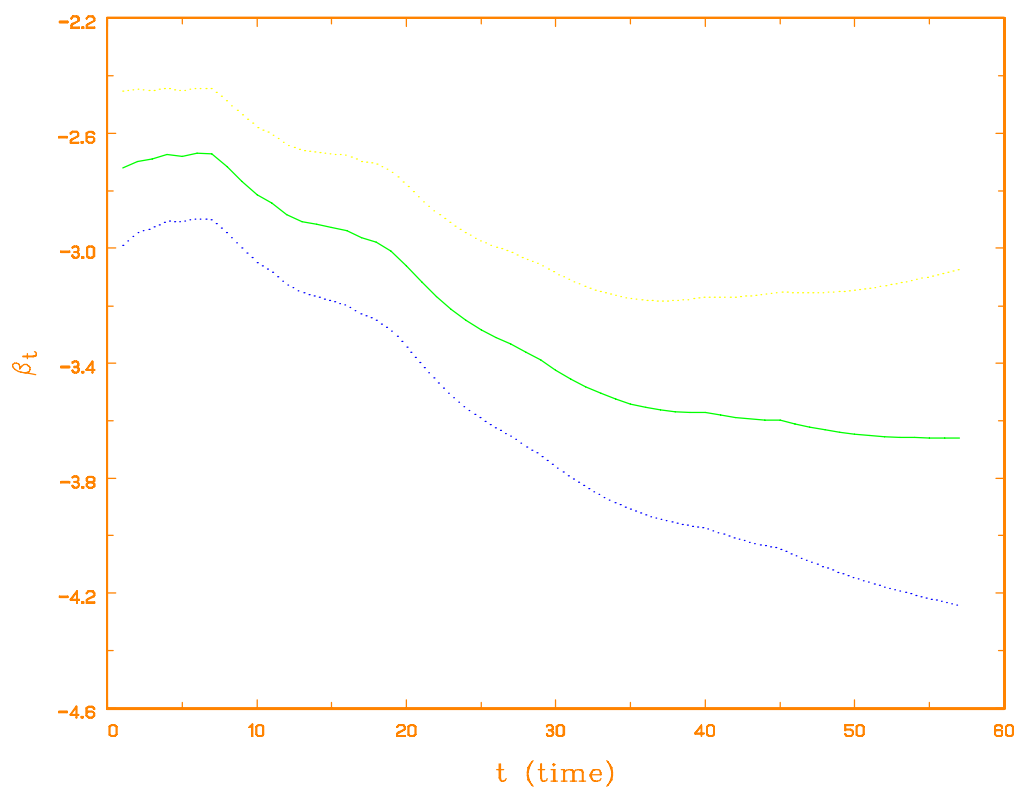


Abbildung 6.1: Geschätzter Baseline-Parameter mit Konfidenzband, 1.Therapiegruppe,
Zeit in Monaten.

In Abbildung 6.1 ist der geschätzte Baseline-Parameter mit Konfidenzband dargestellt. Sie
zeigt, abgesehen von den ersten sieben Monaten, eine fallende Tendenz. Das bedeutet, daß
mit zunehmender Überlebenszeit die Sterberate geringer wird.

Für die zweite Therapiegruppe, die lediglich mit Chemotherapie behandelt wurde, erhalten wir die folgenden Ergebnisse, die in Tabelle 6.2 und Abbildung 6.2 dargestellt sind.

Magenkrebsdaten: Subgruppenanalyse: 2.Therapiegruppe					
$\mathbf{Q}^{(0)}$	$\hat{\mathbf{Q}}$	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.009	0.0091	6	$1 \cdot 10^{-10}$	1	$2 \cdot 10^{-4}$

Tabelle 6.2: Hyperparameterschätzung und Programmablauf von PME.ST.

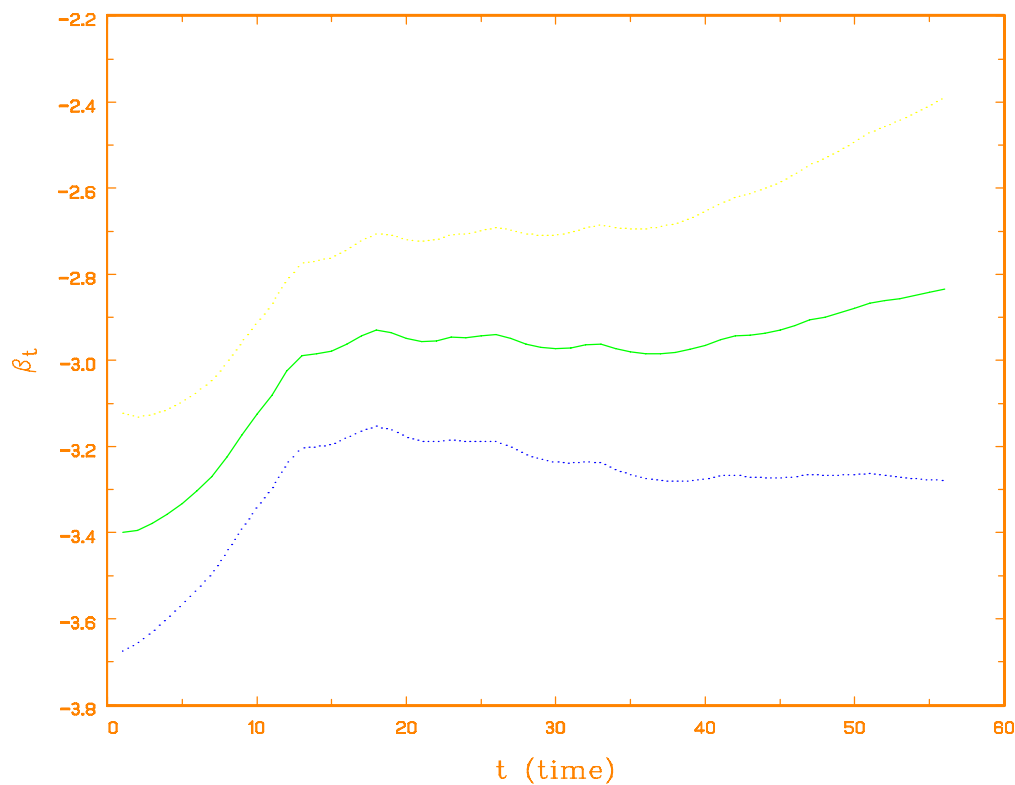


Abbildung 6.2: Geschätzter Baseline-Parameter mit Konfidenzband, 2.Therapiegruppe, Zeit in Monaten.

Vergleicht man den Baseline-Parameter in Abbildung 6.1 mit dem in Abbildung 6.2, so ergibt sich ein entgegengesetzter Verlauf. Um nun zu erfahren, welche der beiden Behandlungsmethoden die bessere ist, plotten wir in Abbildung 6.3 die geschätzten Survivor-Funktionen. Die Schätzungen erhalten wir, indem wir aus den geschätzten Zuständen $\mathbf{a}_{t|\overline{T}}$

mit (6.5) die geschätzten Hazardraten bestimmen und über (2.7) die Werte der Survivor-Funktion.

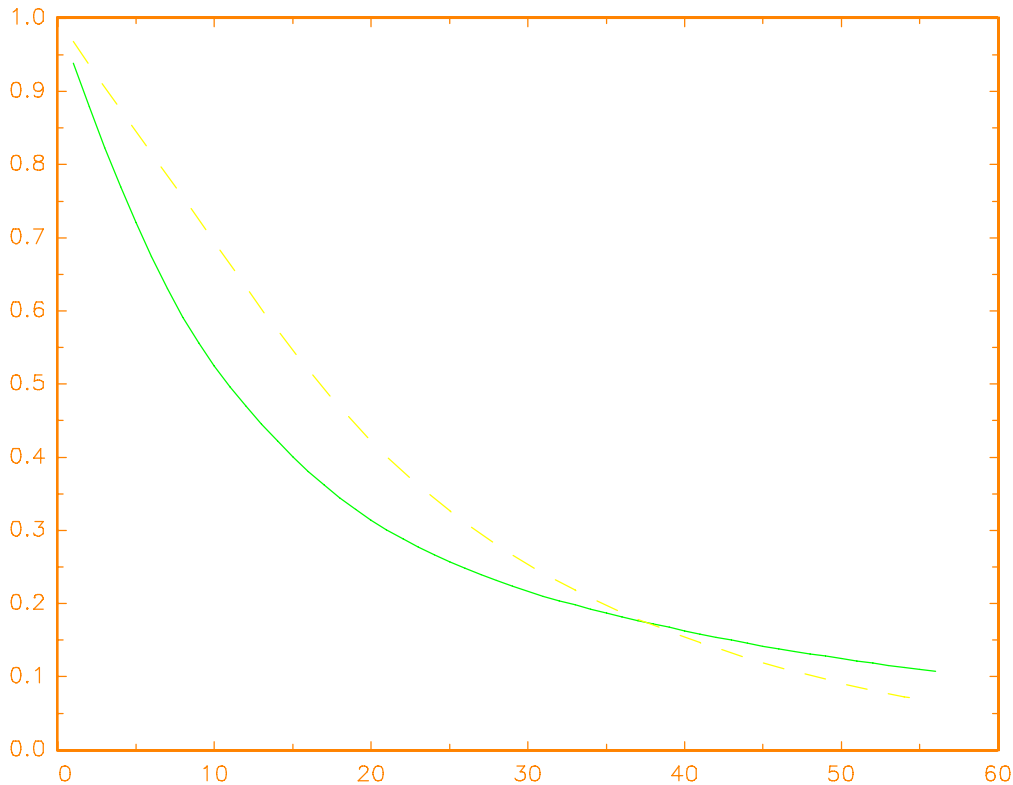


Abbildung 6.3: Geschätzte Survivor-Funktionen, — Chemo- und Strahlentherapie, — — — Chemotherapie, Zeit in Monaten.

Für Überlebenszeiten bis 37 Monate erzielt die Chemotherapie größere Survivor-Funktionswerte und damit Überlebenswahrscheinlichkeiten. Ab 37 Monaten kehrt sich die Situation um, und die kombinierte Therapie ist die „bessere“. Als Ergebnis bleibt festzuhalten: Für kurze Überlebenszeiten ist die Chemotherapie allein die bessere Behandlungsmethode. Ist ein Patient jedoch so robust, die kombinierte Therapie mindestens 37 Monate durchzustehen, dann ist die Chemo- und Strahlentherapie die „bessere“ Behandlungsmethode für ihn.

Gamerman (1991) kommt zu einem ähnlichen Ergebnis, wobei dort der „Change point“ nach circa 1030 Tagen, das entspricht dem 35. Zeitintervall bei Monatseinteilung, auftritt.

Um auch den Therapieeffekt schätzen zu können, haben wir eine Gesamtanalyse vorgenommen. Die Kovariable „Therapiezugehörigkeit“ wurde dabei dummy-kodiert nach

$$\mathbf{x} = \begin{cases} 1, & \text{Chemo- und Strahlentherapie} \\ 0, & \text{Chemotherapie} \end{cases} \quad (6.7)$$

Für das Beobachtungsmodell haben wir wieder ein dynamisches binomiales Logit-Modell gewählt:

$$\lambda(t|\mathbf{x}_U) = \frac{\exp(\eta_{tU})}{1 + \exp(\eta_{tU})}, \quad \eta_{tU} = \tau_t + \mathbf{x}_U \beta_t \quad t = 1, \dots, \bar{T} = 56, U = 1, \dots, \bar{U} = 90. \quad (6.8)$$

τ_t ist dabei der Baseline-Parameter und β_t der Kovariableneffekt. Als Übergangsmodell für die Zustandsvektoren $\alpha_t = (\tau_t, \beta_t)'$ mit $n = \dim(\alpha_t) = 2$ verwenden wir den RW1:

$$\alpha_t = \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(\mathbf{0}_2, Q), \quad t = 1, \dots, \bar{T} = 56. \\ Q \text{ diagonal, Verteilung von } \alpha_0 : \text{diffus.} \quad (6.9)$$

Den algorithmischen Verlauf der Schätzung faßt Tabelle 6.3 zusammen:

Magenkrebsdaten: Gesamtanalyse mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.009	0.008939	6	$1 \cdot 10^{-10}$	1	$4 \cdot 10^{-4}$
0.015	0.015327				

Tabelle 6.3: Hyperparameterschätzung und Programmablauf von PME.ST.

Der geschätzte Baseline-Parameter ist praktisch identisch mit dem Baseline-Parameter in Abbildung 6.2. Den geschätzten, zeitabhängigen Therapieeffekt zeigt die Abbildung 6.4. Zunächst ist der Therapieeffekt signifikant positiv. Mit zunehmender Überlebenszeit nimmt er jedoch ab und schneidet bei circa 21 Monaten das Nullniveau. Das bedeutet: Bis zu einer Überlebenszeit von 21 Monaten hat die kombinierte Therapie einen positiven Einfluß auf die Schätzung der Hazardrate und damit ein höheres Sterberisiko. Ab 21 Monaten kehrt sich die Situation um, wobei das obere Konfidenzband im Bereich des Nullniveaus liegt.

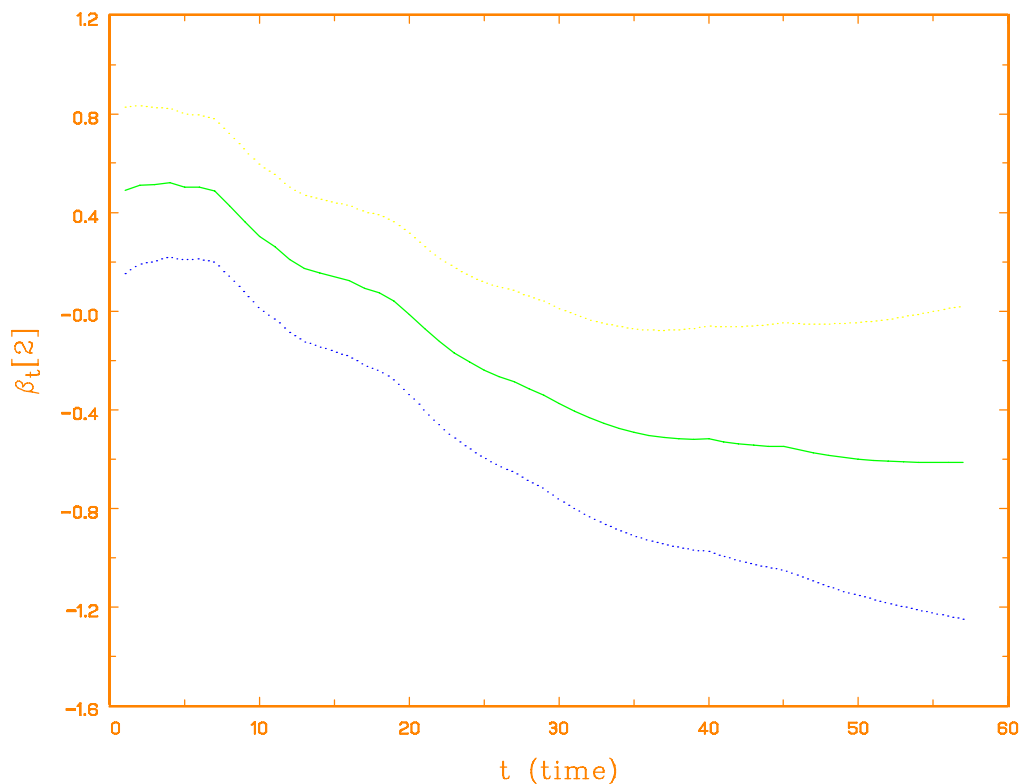


Abbildung 6.4: *Geschätzter Therapieeffekt mit Konfidenzband, Gesamtanalyse mit RWI, Zeit in Monaten.*

In Abbildung 6.5a geben wir die geschätzten Hazardraten für die erste und zweite Behandlungsgruppe an. Die beiden Hazardraten schneiden sich dort, wo der Therapieeffekt die Nullniveaulinie kreuzt: bei 21 Monaten. Die geschätzten Survivor-Funktionen sind in Abbildung 6.5b widergegeben. Die Kurven schneiden sich bei etwa 41 Monaten.

Die Ergebnisse der Gesamtanalyse bestätigen die Resultate der Subgruppenanalyse. Lediglich der „Changepoint“, der Schnitt der beiden Survivor-Kurven, hat sich von 36 Monaten auf 41 nach hinten verschoben.

Die bisher dargestellten Resultate entsprechen den Ergebnissen in Aydemir (1992). Durch die Verwendung einer diffusen Start-Priori-Verteilung im zugrundeliegenden Zustandsraummodell konnte erreicht werden, daß die Konfidenzbänder im Bereich um $t = 0$ nicht unnatürlich eng und zusammengezogen sind.

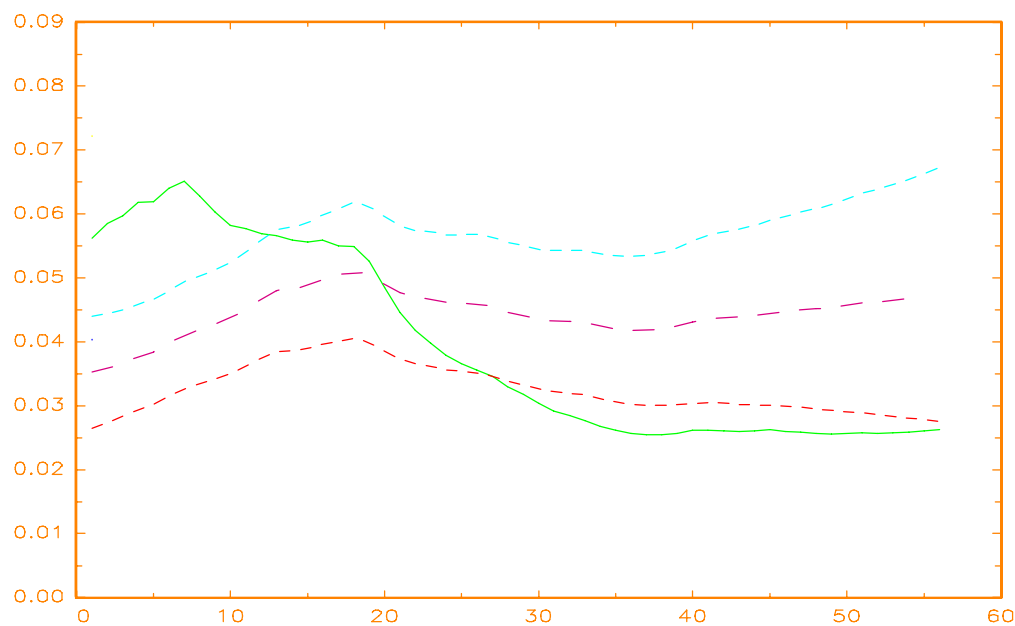


Abbildung 6.5a: Geschätzte Hazardrate für Chemo-und Strahlentherapie ———, geschätzte Hazardrate für Chemotherapie - - - - mit Konfidenzband - - - - , Gesamtanalyse mit RW1, Zeit in Monaten.

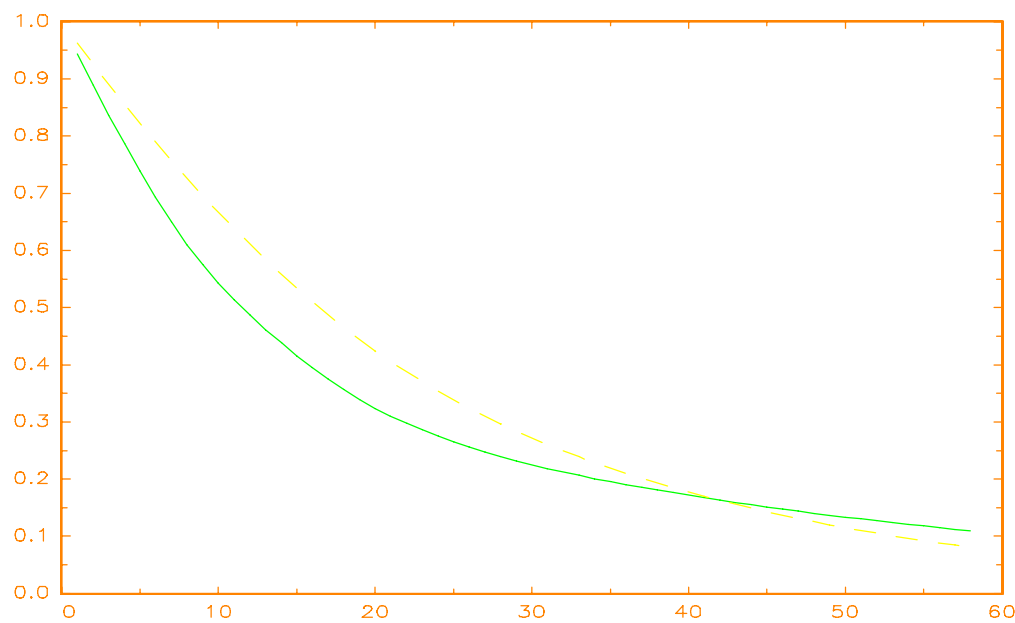


Abbildung 6.5b: Geschätzte Survivor-Funktionen für Chemo-und Strahlentherapie ——— ,für Chemotherapie - - - - Gesamtanalyse mit RW1, Zeit in Monaten.

Um den Einfluß des Übergangsmodells auf die Resultate bei der Schätzung zu erfahren, haben wir die Gesamtanalyse auch durchgeführt mit einem RW2 für den Parameterübergang. Insbesondere soll auch gezeigt werden, daß der EM-Typ Algorithmus und der (D-IGKFG) nicht nur theoretisch, sondern auch in der Praxis mit dem RW2 eingesetzt werden können.

Unser Übergangsmodell hat die Gestalt:

$$\alpha_t = F_t \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(\mathbf{0}_4, Q), \quad t = 1, \dots, \bar{T} = 56.$$

$$Q \text{ diagonal, Verteilung von } \alpha_0 : \text{diffus.} \quad (6.10)$$

mit $\alpha_t = (\tau_t, \tau_{t-1}, \beta_t, \beta_{t-1})'$ und

$$F_t = \begin{pmatrix} 2 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

Über den Programmablauf gibt Tabelle 6.4 Auskunft. Der mit RW2 geschätzte Baseline-Parameter $\hat{\tau}_t$ ist in Abbildung 6.6 und der mit RW2 geschätzte Therapieeffekt $\hat{\beta}_t$ ist in Abbildung 6.7 dargestellt.

Magenkrebsdaten: Gesamtanalyse mit RW2					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.009	0.00838	8	$1 \cdot 10^{-10}$	1	$1 \cdot 10^{-3}$
0	$3 \cdot 10^{-18}$				
0.015	0.01439				
0	$-1 \cdot 10^{-18}$				

Tabelle 6.4: Hyperparameterschätzung und Programmablauf von PME.ST.

Zwei Aspekte sind beim Vergleich der RW1- mit der RW2-Schätzung augenfällig: Die RW2-Schätzung ist glatter, mit weniger Krümmung versehen und gleichzeitig von größerem Ausschlag nach unten und oben. Der Therapieeffekt hat beim RW2 in den ersten Zeitintervallen einen viel größeren Wert als beim RW1, während der Baseline-Parameter einen viel kleineren hat. Umgekehrt verhält es sich gegen Ende des Beobachtungszeitraums.

Eine Erklärung hierfür ist, daß beim RW1 der Strafterm in der penalisierten Loglikelihood der mit Q gewichtete quadratische, euklidische Abstand zwischen α_t und α_{t-1} ist. Hier werden also große Veränderungen der Zustände beim Übergang vom Zeitintervall $(t-1)$ nach t bestraft. Den RW2 hingegen kann man, siehe dazu Fahrmeir und Tutz (1994, Abschnitt 8.1), als zeitdiskretes Analogon zum zeitstetigen kubischen Splineglätter betrachten. Hier werden große Krümmungen im Sinne von integrierten, quadratischen zweiten Ableitungen von $\tau(t)$ beziehungsweise $\beta(t)$ als Funktion von t bestraft.

In Abbildung 6.8 geben wir noch die zugehörigen geschätzten Hazardraten an.

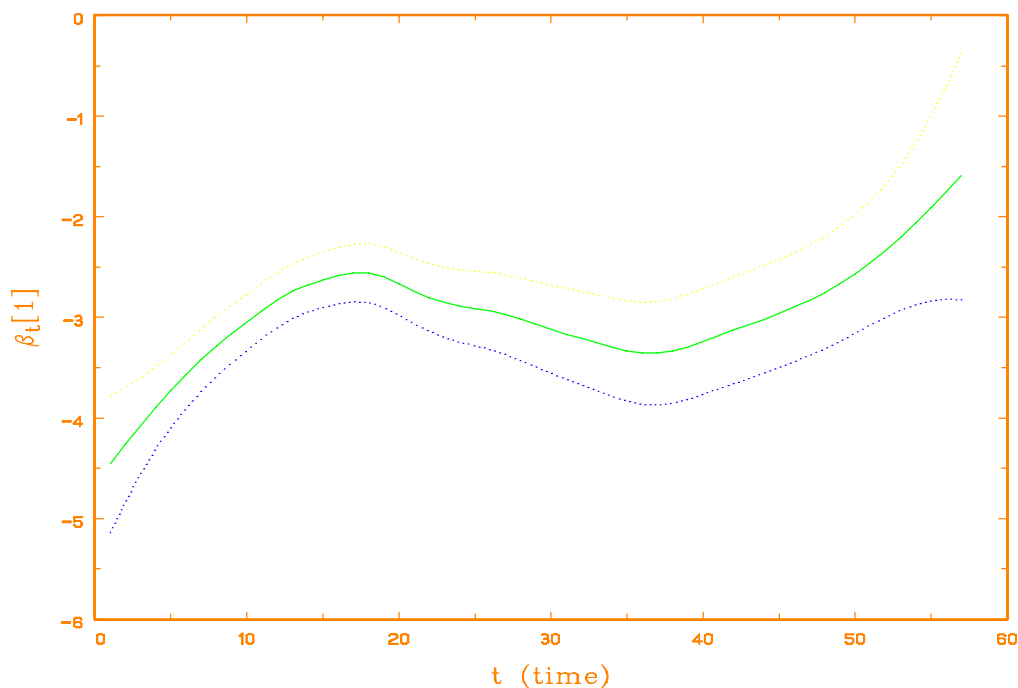


Abbildung 6.6: Geschätzter Baseline-Parameter mit Konfidenzband, Gesamtanalyse mit RW2, Zeit in Monaten.

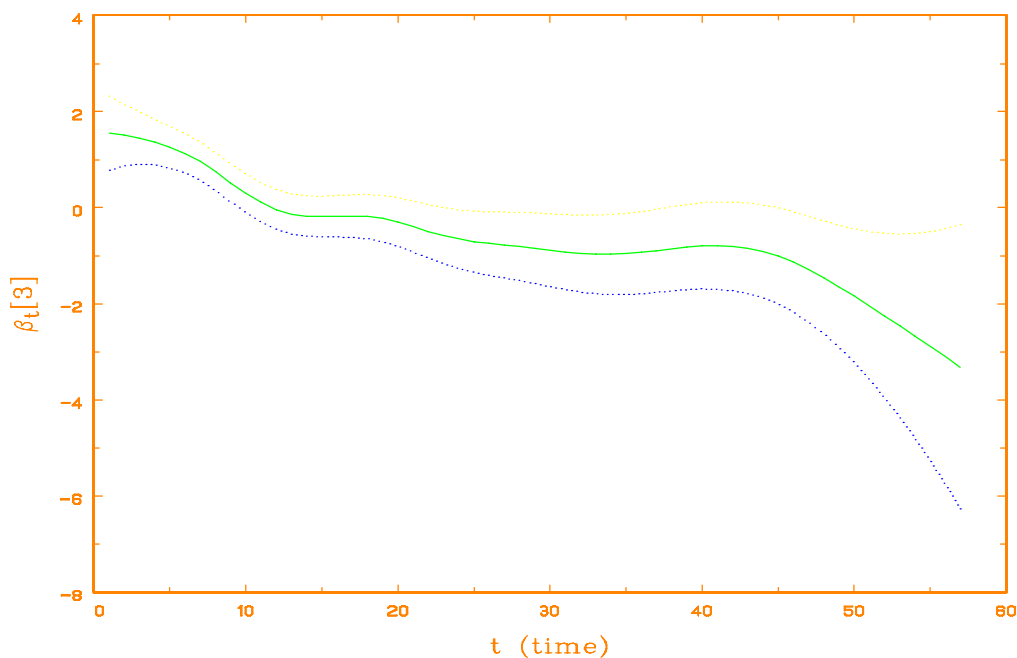


Abbildung 6.7: Geschätzter Therapieeffekt mit Konfidenzband, Gesamtanalyse mit RW2, Zeit in Monaten.

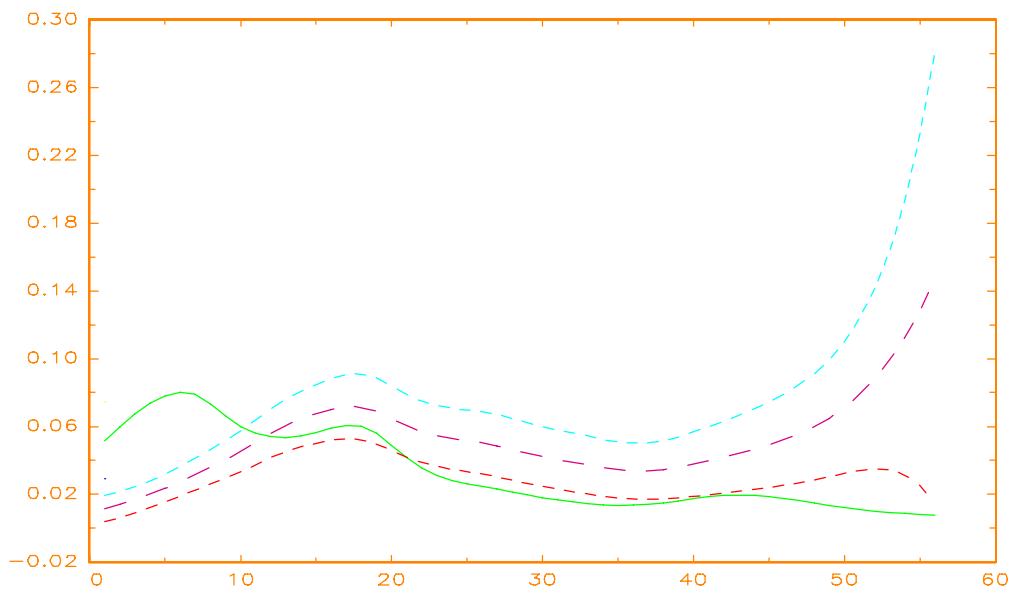


Abbildung 6.8: Geschätzte Hazardrate für Chemo-und Strahlentherapie —, geschätzte Hazardrate für Chemotherapie - - - mit Konfidenzband - - - -, Gesamtanalyse mit RW2, Zeit in Monaten.

6.1.2 Piecewise exponential Modell

Im folgenden wenden wir das Piecewise exponential Modell (3.23), (3.7) zur Analyse der Magenkrebsdaten an. Als Übergangsmodell verwenden wir dabei stets den RW1. Wesentliches Unterscheidungsmerkmal der verschiedenen Modellansätze sind die verschiedenen Zeitgrids b , die wir benutzen. Als Zeiteinheit verwenden wir in diesem Teilabschnitt immer Tage.

Das Zeitgrid b sei gegeben durch Zeitintervalle von 30 Tagen Länge, also $b = \{0,30,60,...,1740\}$. Diese Einteilung entspricht den Monatsdaten in Teilabschnitt 6.1.1, jedoch in einer anderen Skalierung. Für die Fehlerkovarianzmatrix im Übergangsmodell haben wir $cov(\xi_t) = w_t \cdot Q_t$ benützt. Die nachfolgende Tabelle 6.5 gibt Auskunft über den Programmablauf:

Magenkrebsdaten: Piecewise exponential Modell mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.0001	0.0001	13	$1 \cdot 10^{-10}$	1	$1 \cdot 10^{-5}$
0.0005	0.000508				

Tabelle 6.5: Hyperparameterschätzung und Programmablauf von *PME.ST*.

In Abbildung 6.9 ist der geschätzte Baseline-Parameter widergegeben und in Abbildung 6.10 der geschätzte Therapieeffekt.

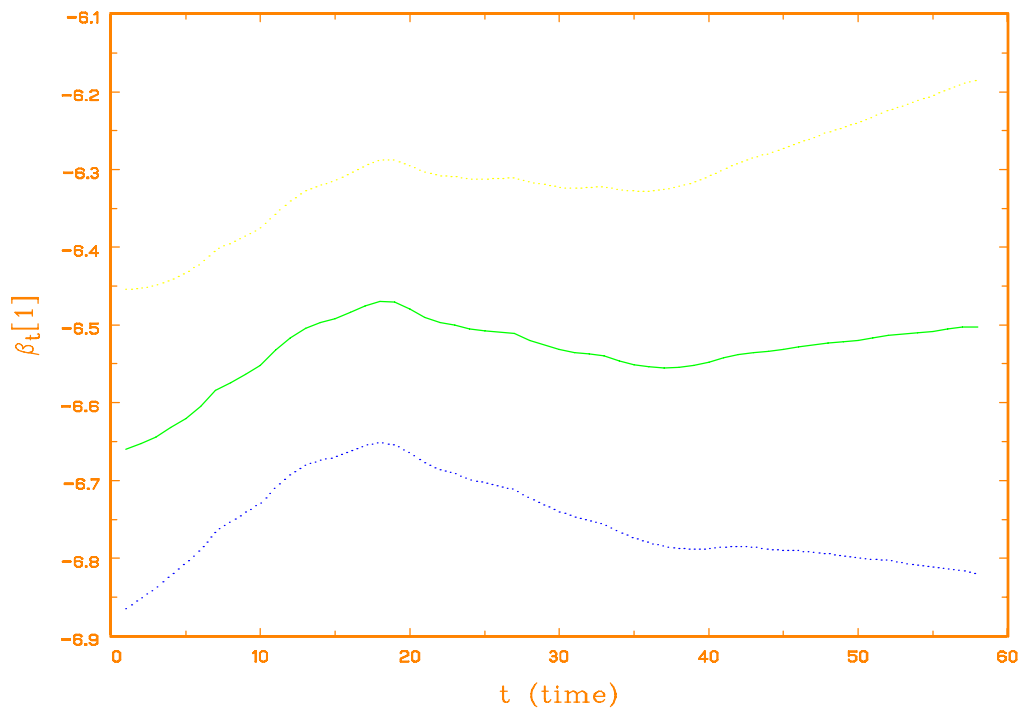


Abbildung 6.9: Geschätzter Baseline-Parameter mit Konfidenzband, Piecewise exponential Modell mit RW1. t gibt die Zeitintervallnummer an.

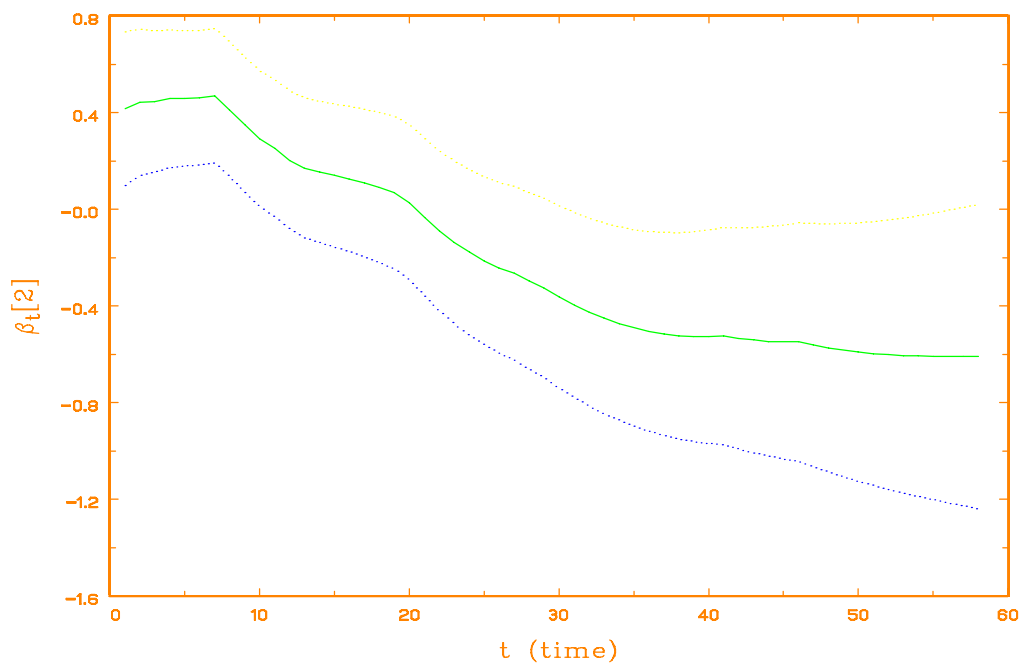


Abbildung 6.10: Geschätzter Therapieeffekt mit Konfidenzband, Piecewise exponential Modell mit RW1. t gibt die Zeitintervallnummer an.

Die Schätzung des Therapieeffektes ist praktisch identisch mit der Schätzung im Rahmen der Gesamtanalyse mit RW1 in Teilabschnitt 6.1.1. Auch der Baseline-Parameter hat die gleiche Gestalt. Sie ist lediglich um etwa 3.5 Einheiten parallel nach unten verschoben. Wegen (4.41) hat dies jedoch keine Auswirkung auf die Schätzung der Survivor-Funktion. Sie ist identisch mit der Darstellung in Abbildung 6.5b.

Benutzt man statt der obigen die in (4.39) angegebene Spezifizierung der Fehlerkovarianzmatrix, so erhält man bei der Wahl von $\mathbf{Q}^{(0)} = \text{diag}(0.09, 0.45)$ die gleichen Ergebnisse. Dies ist darauf zurückzuführen, daß bei obigem Zeitgrid die Aussage in (4.40) gilt.

Gamerman (1991) schlägt vor, die beobachteten Todes- beziehungsweise Abgangstage als Zeitgrid zu verwenden. Damit ist $\mathbf{b} = \{0, 1, 17, 42, 44, 48, 60, 63, 72, 74, 95, 103, 105, 108, 122, 125, 144, 167, 170, 182, 183, 185, 193, 195, 197, 208, 216, 234, 235, 250, 254, 262, 301, 307, 315, 342, 354, 356, 358, 380, 383, 388, 394, 401, 408, 445, 460, 464, 484, 489, 499, 523, 524, 528, 535, 542, 562, 567, 569, 577, 580, 675, 676, 748, 778, 786, 795, 797, 855, 955, 968, 977, 1174, 1214, 1232, 1245, 1271, 1366, 1420, 1455, 1460, 1516, 1551, 1585, 1622, 1626, 1690, 1694, 1736\}$.

Mit der Spezifikation $\text{cov}(\xi_t) = \mathbf{w}_t \cdot \mathbf{Q}_t$ erhalten wir die folgenden Ergebnisse, die in Tabelle 6.6 und den Abbildungen 6.11 bis 6.13 festgehalten sind.

Magenkrebsdaten: Piecewise exponential Modell mit RW1					
$\text{diag}(\mathbf{Q}^{(0)})$	$\text{diag}(\hat{\mathbf{Q}})$	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.00003	0.00003	13	$1 \cdot 10^{-10}$	1	$3 \cdot 10^{-6}$
0.00015	0.0001527				

Tabelle 6.6: Hyperparameterschätzung und Programmablauf von PME.ST.

Sowohl der Baseline-Parameter in Abbildung 6.11 als auch der Therapieeffekt in Abbildung 6.12 nehmen am Anfang einen glatten Verlauf, während gegen Ende, wo die beobachteten Abgangszeiten weit auseinander liegen, Sprünge und größere Rauheit zu beobachten sind. Zur Verdeutlichung haben wir hier als Zeiteinheit auf der Abszisse die Zeitintervallnummer gewählt. Der Therapieeffekt, zu Beginn signifikant positiv, schneidet beim 62. Zeitintervall

das Nullniveau. Der „Changepoint“ bei den Survivor-Kurven in Abbildung 6.13 liegt bei etwa $t = 1550$, das ist im 81. Zeitintervall. Damit hat sich der Schnittpunkt der Survivor-Funktionen im Vergleich zur Monatsanalyse ($t = 1230$) in Teilabschnitt 6.1.1 weiter nach hinten verlagert.

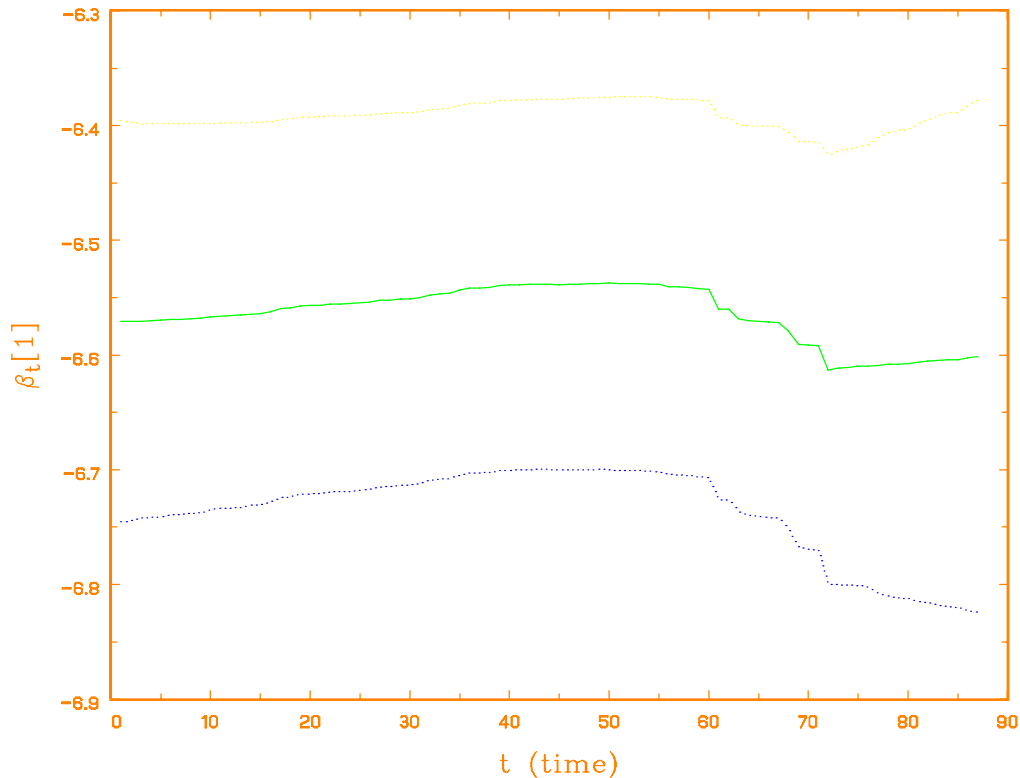


Abbildung 6.11: Geschätzter Baseline-Parameter mit Konfidenzband, Piecewise exponential Modell mit RW1, beobachtete Abgangszeiten als Zeitgrid. t gibt die Zeitintervallnummer an.

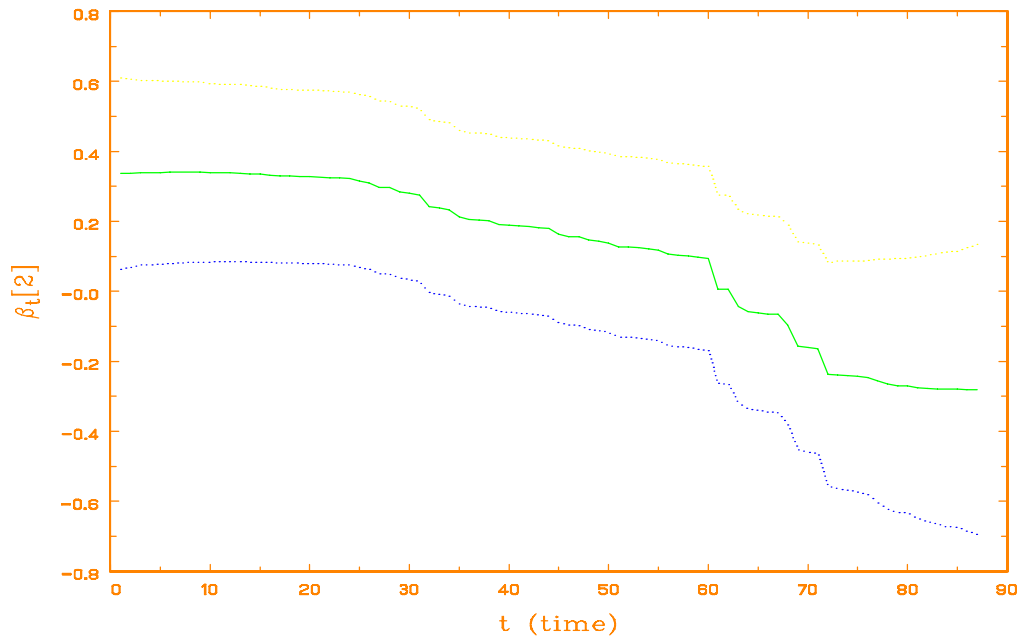


Abbildung 6.12: Geschätzter Therapieeffekt mit Konfidenzband, Piecewise exponential Modell mit RW1, beobachtete Abgangszeiten als Zeitgrid. t gibt die Zeitintervallnummer an.

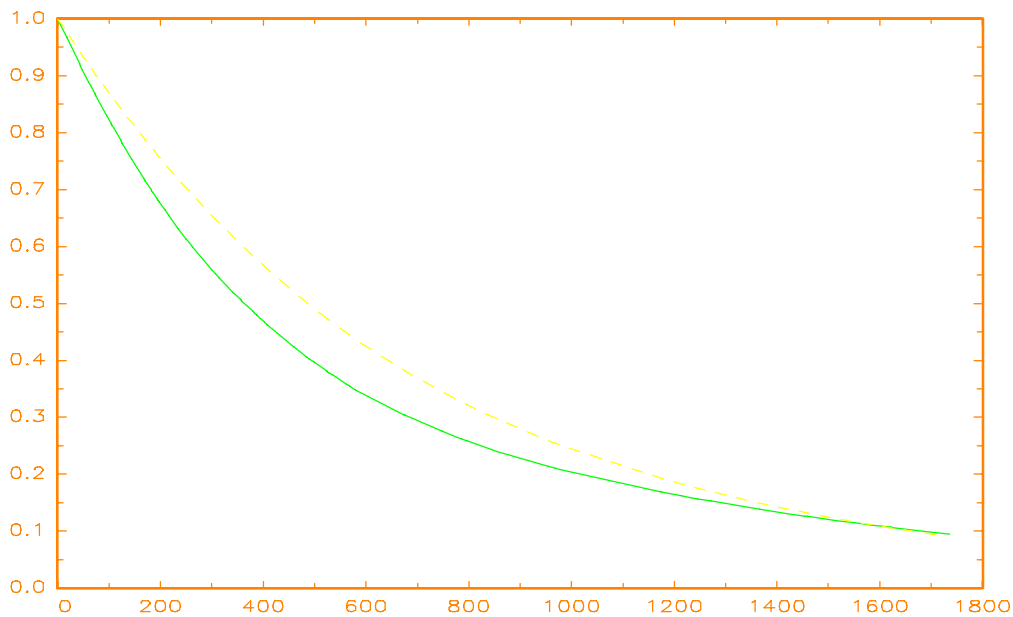


Abbildung 6.13: Geschätzte Survivor-Funktionen für Chemo- und Strahlentherapie ———, für Chemotherapie — — — —. Piecewise exponential Modell mit RW1, beobachtete Abgangszeiten als Zeitgrid. Abszisse: Zeit in Tagen.

Benutzt man bei gleichem Zeitgrid für die Fehlerkovarianzmatrix im Übergangsmodell die Spezifikation (4.39), so bekommt man wir die folgenden Resultate, die in Tabelle 6.7 und den Abbildungen 6.14 und 6.15 dargestellt sind.

Magenkrebsdaten: Piecewise exponential Modell mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.003	0.003	13	$1 \cdot 10^{-10}$	1	$3 \cdot 10^{-4}$
0.015	0.015221				

Tabelle 6.7: Hyperparameterschätzung und Programmablauf von PME.ST.

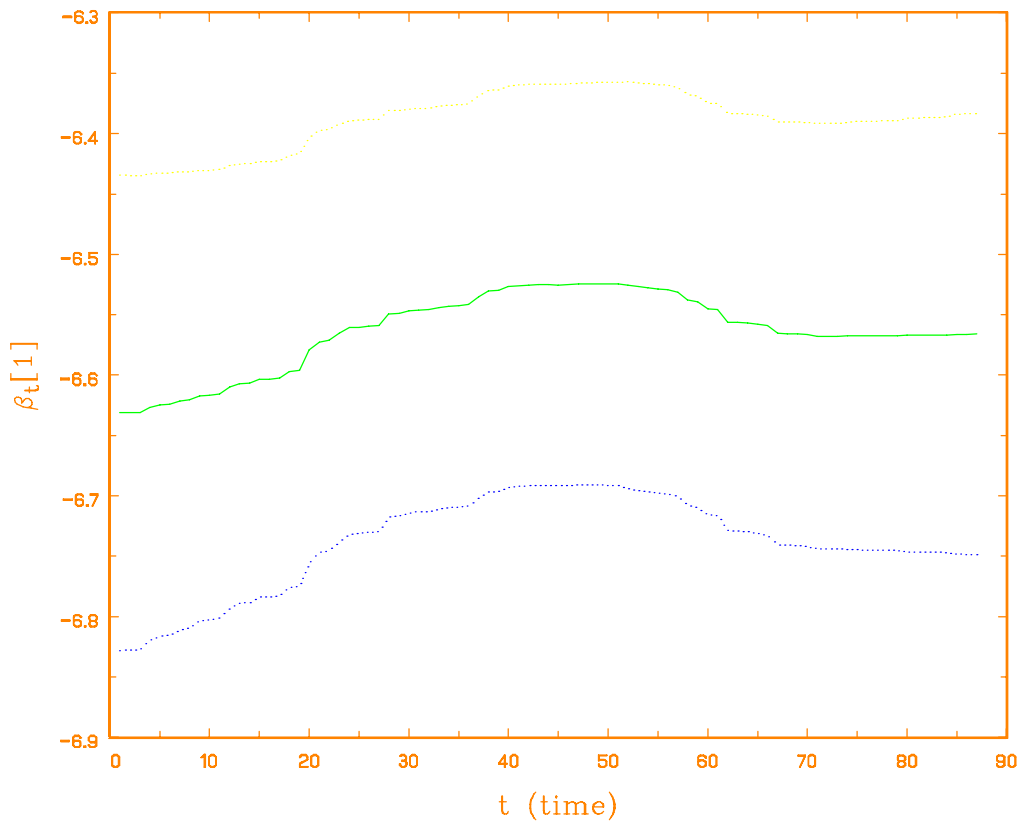


Abbildung 6.14: Geschätzter Baseline-Parameter mit Konfidenzband, Piecewise exponential Modell mit RW1, beobachtete Abgangszeiten als Zeitgrid. t gibt die Zeitintervallnummer an.

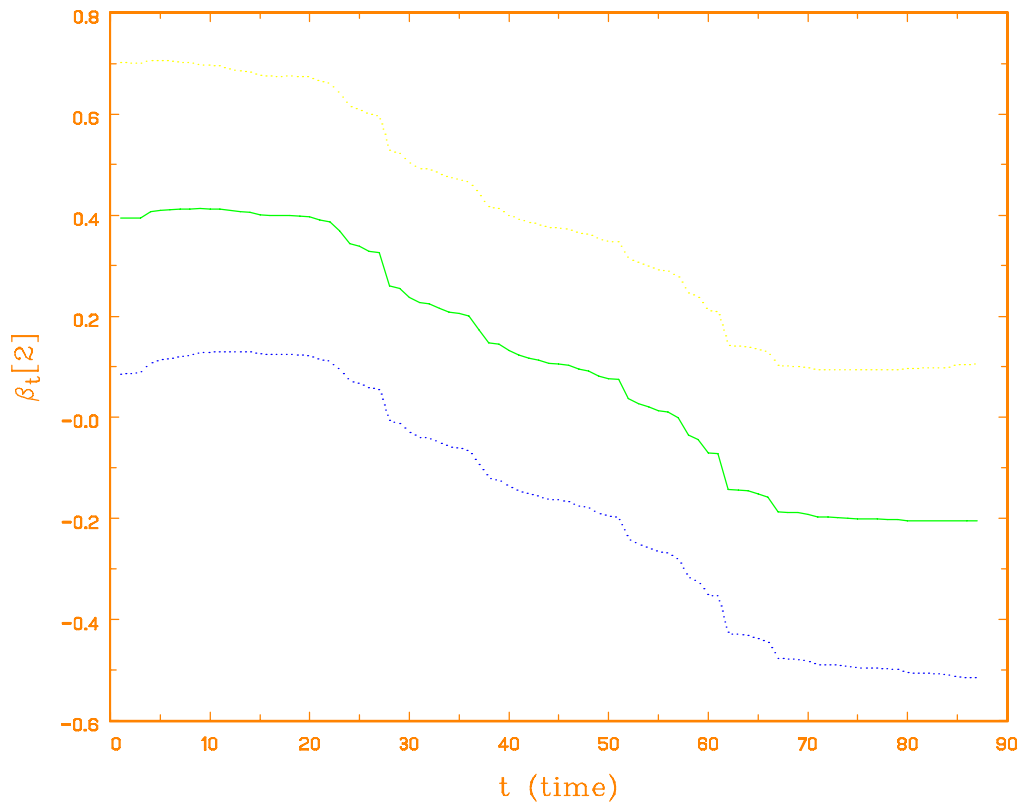


Abbildung 6.15: Geschätzter Therapieeffekt mit Konfidenzband, Piecewise exponential Modell mit RW1, beobachtete Abgangszeiten als Zeitgrid. t gibt die Zeitintervallnummer an.

Der geschätzte Baselineeffekt in Abbildung 6.14 und auch der geschätzte Therapieeffekt in Abbildung 6.15 sind von der Gestalt und vom Niveau her vergleichbar mit den Resultaten in Abbildung 6.11 respektive 6.12. Bei der Spezifikation wie in (4.39) ergibt sich jedoch eine über alle Zeitintervalle hinweg gleichmäßigere Rauheit beziehungsweise Glattheit. Dieses Ergebnis stimmt überein mit unseren Überlegungen zur Spezifikation der Fehlerkovarianzmatrix des Übergangsmodells in Abschnitt 4.1. Die geschätzten Survivor-Funktionen, die man hier erhält, sind visuell nicht von der Abbildung 6.13 unterscheidbar.

Gamerman (1991) schlägt zur Analyse ein weiteres Zeitgrid $b = \{0, 20, 40, 60, \dots, 180, 200, 250, 300, \dots, 550, 600, 700, 800, \dots, 1700, 1800\}$ vor, bei dem am Anfang, wo viel Dateninformation vorhanden ist, die Zeitintervalle relativ kurz, und am Ende relativ lang sind. Mit

der Spezifikation der Fehlerkovarianzmatrix $\text{cov}(\xi_t) = \mathbf{w}_t \cdot \mathbf{Q}_t$ erhalten wir die folgenden Schätzungen, deren algorithmische Bestimmung in Tabelle 6.8 zusammengefaßt ist.

Magenkrebsdaten: Piecewise exponential Modell mit RW1					
$\text{diag}(\mathbf{Q}^{(0)})$	$\text{diag}(\hat{\mathbf{Q}})$	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.00003	0.00003014	13	$1 \cdot 10^{-10}$	1	$2 \cdot 10^{-5}$
0.0007	0.0007115				

Tabelle 6.8: Hyperparameterschätzung und Programmablauf von *PME.ST*.

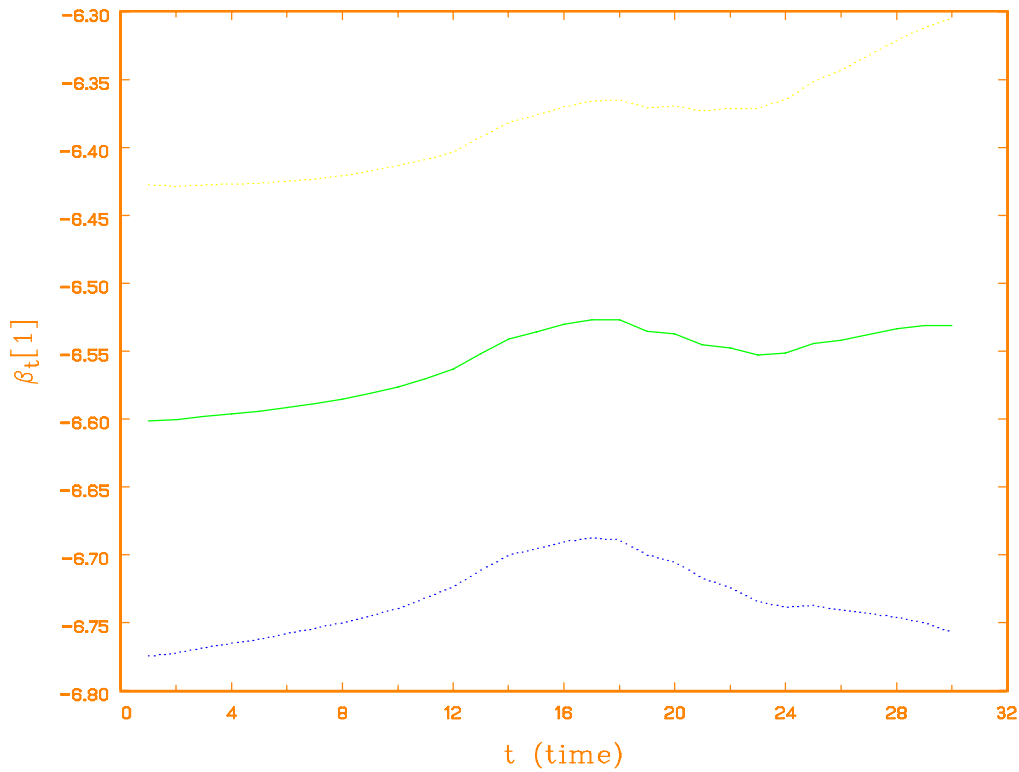


Abbildung 6.16: Geschätzter Baseline-Parameter mit Konfidenzband, Piecewise exponential Modell mit RW1, unterschiedliche Zeitintervalllängen als Zeitgrid. *t* gibt die Zeitintervallnummer an.

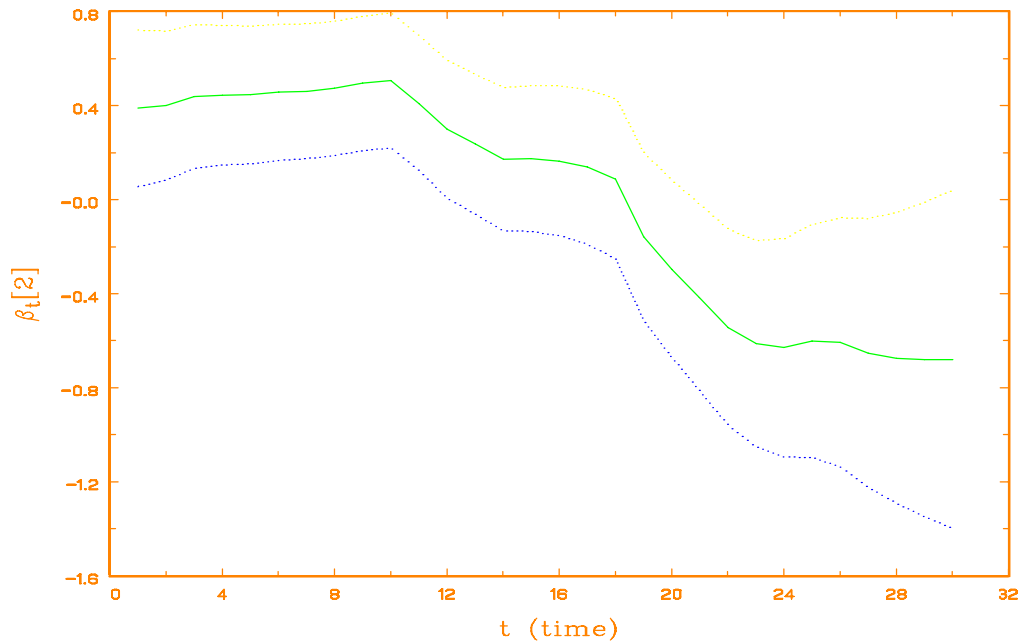


Abbildung 6.17: Geschätzter Therapieeffekt mit Konfidenzband, Piecewise exponential Modell mit RW1, unterschiedliche Zeitintervalllängen als Zeitgrid. t gibt die Zeitintervallnummer an.

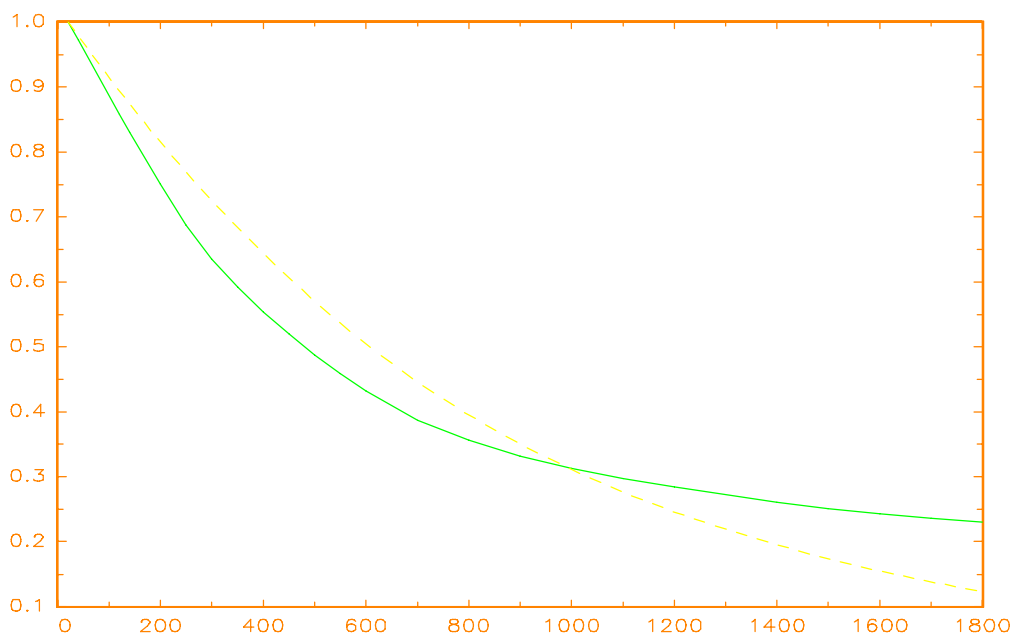


Abbildung 6.18: Geschätzte Survivor-Funktionen für Chemo- und Strahlentherapie ———, für Chemotherapie - - - -. Piecewise exponential Modell mit RW1, unterschiedliche Zeitintervalllängen als Zeitgrid. Abszisse: Zeit in Tagen.

Der geschätzte Baseline-Parameter in Abbildung 6.16 nimmt einen glatten Verlauf auf einem Niveau, das vergleichbar ist mit den bisherigen Analysen. Am Verlauf des Therapieeffektes in Abbildung 6.17 erkennt man deutlich die Änderung der Zeitintervallängen nach 200 beziehungsweise 600 Tagen an den Knicken im 10. und 18. Zeitintervall. Die geschätzten Survivor-Funktionen in Abbildung 6.18 mit ihrem Schnittpunkt bei etwa 1000 Tagen stimmen sehr gut mit dem Resultat in Gamerman (1991) überein, der für seine Schätzung einen konjugierten Priori-Ansatz zugrunde gelegt hat.

Verwenden wir für die Spezifikation der Fehlerkovarianzmatrix den Ansatz in (4.39), so nehmen die Schätzungen des Baselineeffekts in Abbildung 6.19 und die des Therapieeffekts in Abbildung 6.20 wieder einen im Vergleich zu den Abbildungen 6.16 und 6.17 gleichmäßig glatten Verlauf. Insbesondere sind beim Therapieeffekt in Abbildung 6.20 keine „Knicke“ in den Zeitintervallen 10 und 18 vorhanden. Der Programmablauf ist in Tabelle 6.9 zusammengefaßt. Die geschätzten Survivor-Kurven für die beiden Therapiegruppen sind visuell nicht von den Graphen in Abbildung 6.18 unterscheidbar. Wir erhalten also bezüglich der Survivor-Funktion wieder das gleiche Ergebnis wie bei der Spezifikation $cov(\xi_t) = w_t \cdot Q_t$.

Magenkrebsdaten: Piecewise exponential Modell mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.03	0.03013	26	$1 \cdot 10^{-10}$	2	$2 \cdot 10^{-2}$
0.7	0.72				

Tabelle 6.9: Hyperparameterschätzung und Programmablauf von PME.ST.

FS-Iter. in Tabelle 6.9 gibt die Zahl der insgesamt ausgeführten Fisher-Scoring Iterationen an.

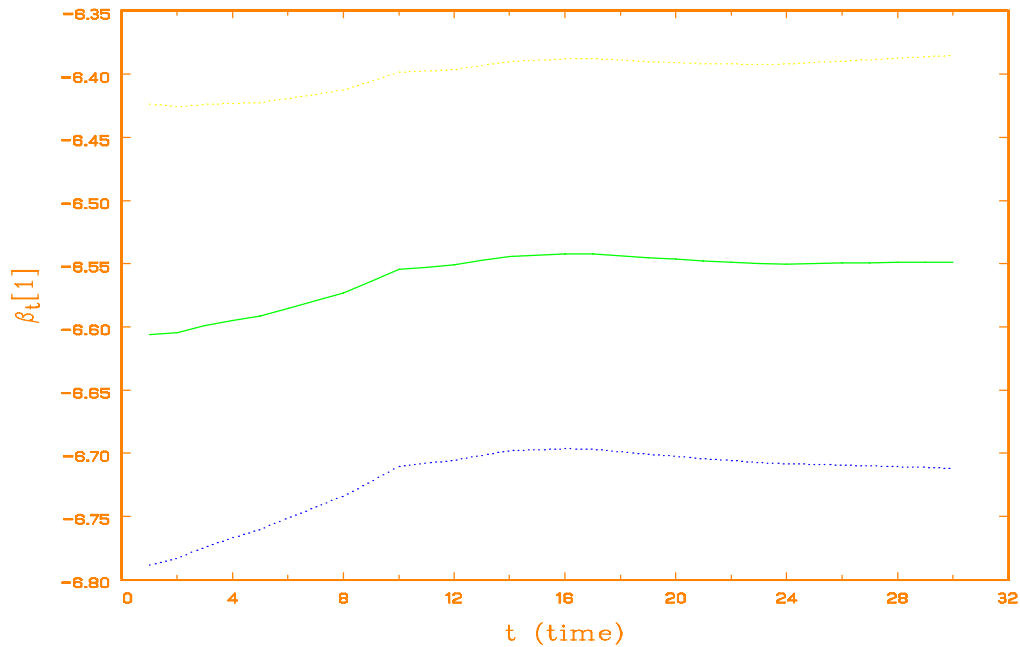


Abbildung 6.19: Geschätzter Baseline-Parameter mit Konfidenzband, Piecewise exponential Modell mit RW1, unterschiedliche Zeitintervalllängen als Zeitgrid. t gibt die Zeitintervallnummer an.

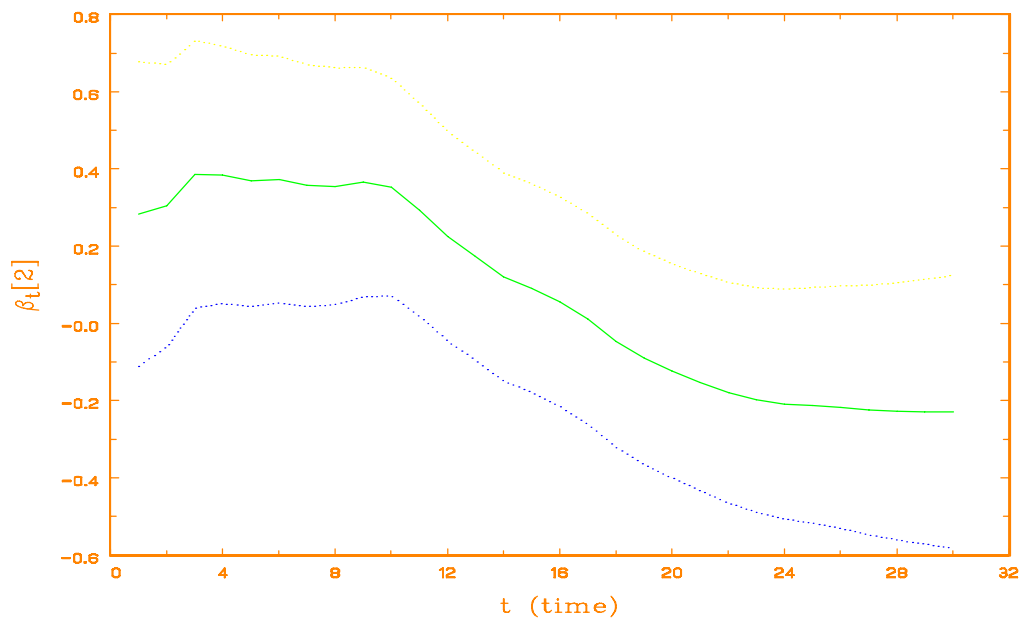


Abbildung 6.20: Geschätzter Therapieeffekt mit Konfidenzband, Piecewise exponential Modell mit RW1, unterschiedliche Zeitintervalllängen als Zeitgrid. t gibt die Zeitintervallnummer an.

6.2 Efron-Daten

Die Efron-Daten wollen wir mit zwei verschiedenen Modellen analysieren: dem dynamischen binomialen Logit-Modell und dem dynamischen gruppierten Cox-Modell. Zur Untersuchung wurden die im Datensatz in Tagen angegebenen Verweildauern wie in Efron (1988) zu Monaten gruppiert. Einen Monat entspricht dabei 30,438 Tage.

Die Kovariable Therapiezugehörigkeit kodieren wir in Dummy-Form:

$$\mathbf{x}_U = \begin{cases} 1 & \text{Patient U erhielt Strahlentherapie} \\ 0 & \text{Patient U erhielt Chemo – und Strahlentherapie} \end{cases} \quad (6.11)$$

Das dynamische binomiale Logit-Modell ist gegeben durch

$$\lambda(t|\mathbf{x}_U) = \frac{\exp(\eta_{tU})}{1 + \exp(\eta_{tU})}, \quad \eta_{tU} = \tau_t + \mathbf{x}_U \beta_t \quad t = 1, \dots, \bar{T} = 76, U = 1, \dots, \bar{U} = 96. \quad (6.12)$$

τ_t ist dabei der Baseline-Parameter und β_t der Therapie- beziehungsweise Kovariablen-effekt. β_t gibt den Einfluß an, den die Zugehörigkeit zur Strahlentherapiegruppe im Bezug auf die Gruppe mit der kombinierten Therapie hat. Als Übergangsmodell für die Zustandsvektoren $\alpha_t = (\tau_t, \beta_t)'$ mit $n = \dim(\alpha_t) = 2$ verwenden wir in diesem Teilabschnitt den RW1:

$$\alpha_t = \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(\mathbf{0}_2, Q), \quad t = 1, \dots, \bar{T} = 76. \\ Q \text{ diagonal, Verteilung von } \alpha_0 : \text{diffus.} \quad (6.13)$$

Der algorithmische Ablauf der Schätzung ist in Tabelle 6.10 zusammengefaßt.

Efron-Daten: Dynamisches binomiales Logit-Modell mit RW1					
$\text{diag}(Q^{(0)})$	$\text{diag}(\hat{Q})$	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.095	0.105	63	$1 \cdot 10^{-10}$	9	$1 \cdot 10^{-3}$
0.005	0.0049				

Tabelle 6.10: Hyperparameterschätzung und Programmablauf von PME.ST.

In Abbildung 6.21 ist der geschätzte Baseline-Parameter mit punktwisen Konfidenzbändern geplottet.

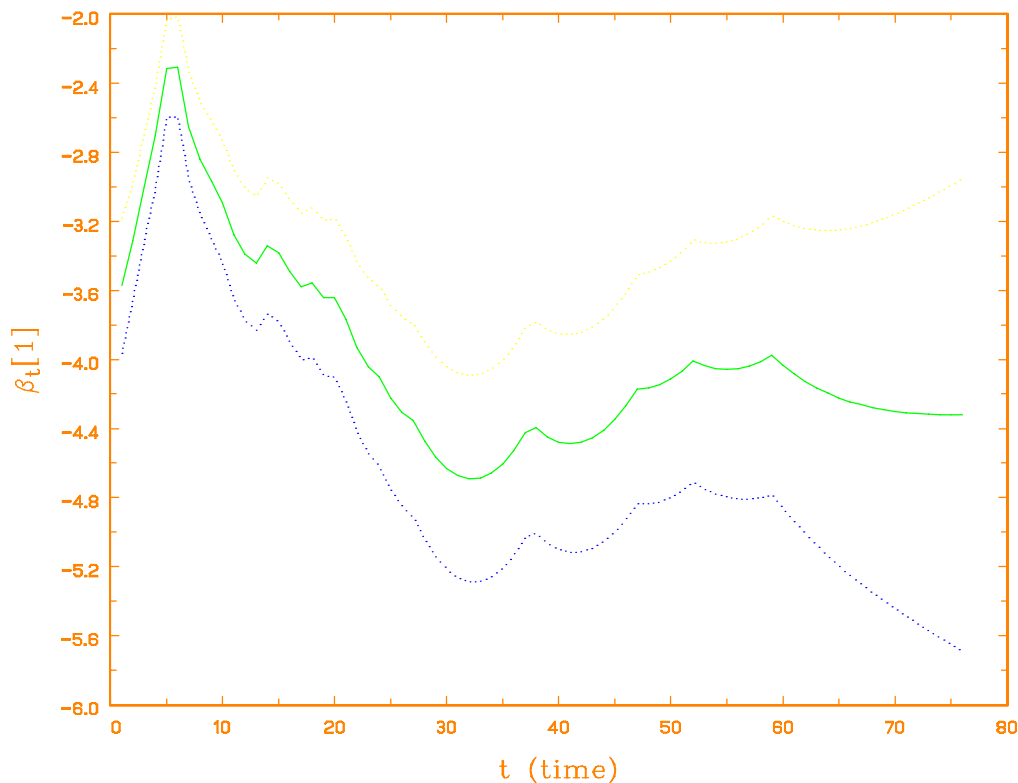


Abbildung 6.21: Geschätzter Baseline-Parameter mit Konfidenzband, dynamisches binomiales Logit-Modell mit RW1. Zeit t in Monaten.

Den geschätzten Therapieeffekt für dieses Modell haben wir bereits in Abbildung 5.2 in Abschnitt 5.2 angegeben.

In einer zweiten Analyse wurde das dynamische gruppierte Cox-Modell nach (3.15b) zusammen mit dem RW1 benutzt:

$$\lambda(t|x_U) = 1 - \exp\{-\exp(\eta_{tU})\}, \quad \eta_{tU} = \tau_t + \mathbf{x}_U \beta_t \quad t = 1, \dots, \bar{T} = 76, U = 1, \dots, \bar{U} = 96. \quad (6.14)$$

Der geschätzte Baseline-Parameter ist identisch mit den in Abbildung 6.21 dargestellten Werten. Deshalb zeigen wir nur noch den Programmablauf in Tabelle 6.11 und die Schätzung des Therapieeffektes in Abbildung 6.22.

Efron-Daten: Dynamisches gruppiertes Cox-Modell mit RW1					
$\text{diag}(\mathbf{Q}^{(0)})$	$\text{diag}(\hat{\mathbf{Q}})$	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.095	0.0991	36	$1 \cdot 10^{-10}$	4	$1 \cdot 10^{-3}$
0.005	0.00498				

Tabelle 6.11: Hyperparameterschätzung und Programmablauf von PME.ST.

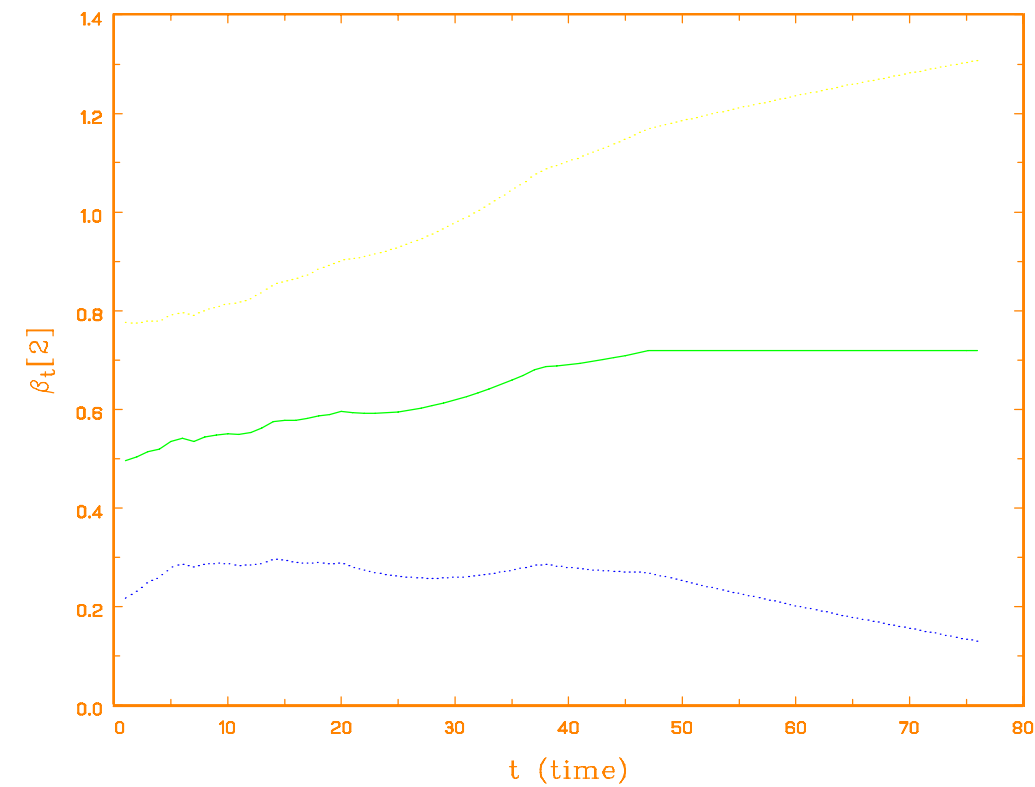


Abbildung 6.22: Geschätzter Therapieeffekt mit Konfidenzband, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten

Zwischen dem geschätzten Therapieeffekt in Abbildung 6.22 und der Schätzung in Abbildung 5.2 gibt es nur minimale Unterschiede. In beiden Fällen liegt der Effekt für den gesamten Beobachtungszeitraum, $t = 1, \dots, 76$, signifikant über dem Nullniveau und zeigt kaum eine zeitliche Variation. Ab dem Zeitintervall $t = 48$ ist er völlig konstant. Der Grund dafür ist, daß die Verweildauerzeiten für Patienten aus der Strahlentherapiegruppe weniger als 48 Monate betragen und somit ab dem 48. Monat kein Patient aus dieser Gruppe mehr

unter Risiko steht. Der Design-Vektor hat also ab $t = 48$ für alle noch in der Risikomenge befindlichen Untersuchungseinheiten den Wert $\mathbf{z}_U = (1, 0)'$. Anhand der penalisierten Loglikelihood (4.37) und der Modellspezifikation in (6.14) erkennt man damit, daß der Posteriori-Modus-Schätzer für β_t , $t = 48, \dots, 76$ nicht eindeutig ist. Für die Optimalität der penalisierten Loglikelihood kann β_t , $t = 48, \dots, 76$ sogar jeden beliebigen Wert annehmen. Die Kalman-Filter-Formeln sind so aufgebaut, daß $\beta_{t|76} = \beta_{47|76}$ für $t = 48, \dots, 76$ gilt.

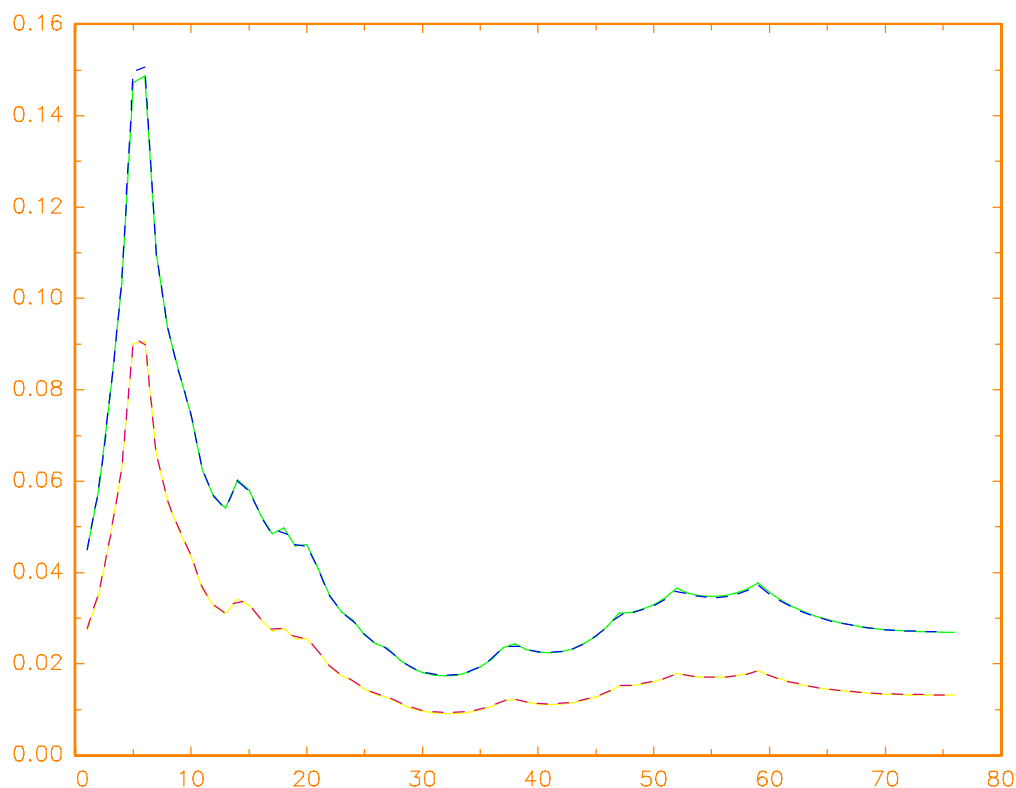


Abbildung 6.23: Geschätzte Hazardraten: — dynamisches binomiales Logit-Modell, — — — dynamisches gruppiertes Cox-Modell, mit RW1. Obere Kurve: Strahlentherapiegruppe, untere Kurve: Chemo- und Strahlentherapie. Abszisse: Zeit t in Monaten.

Die Abbildung (6.23) zeigt die für die Modelle (6.12) und (6.14) geschätzten Hazardraten. Die Rate für die Strahlentherapiegruppe ist, wie es der Therapieeffekt in den Abbildungen 5.2 und 6.22 anzeigt, während des gesamten Beobachtungszeitraumes über der Schätzung

für die kombinierte Therapie. Damit weist sich die Chemo- und Strahlentherapie als die bezüglich Überlebenszeit bessere aus. Zwischen dem dynamischen binomialen Logit-Modell und den dynamischen gruppierten Cox-Modell gibt es bei diesem Datensatz für die Hazardratenschätzung praktisch keinen Unterschied. Die geschätzten Hazardraten in Abbildung (6.23) stimmen auch sehr gut mit den Ergebnissen in Efron (1988) überein. Dort wird ein „partial logistic regression“-Ansatz mit zeitkonstantem Therapieeffekt verwendet. Die gute Übereinstimmung läßt sich dadurch erklären, daß der dynamisch geschätzte Therapieeffekt in den Abbildungen 5.2 und 6.22 wenig Zeitvariation zeigt.

6.3 Veteran's Administration Lungenkrebsdaten

Zur Analyse dieses Datensatzes aus Teilabschnitt 1.1.3 haben wir die Verweildauern, die in Tagen angegeben waren, wegen der geringen Fallzahlen gegen Ende des Beobachtungszeitraumes zu Monaten gruppiert. Die Abbildung 6.24 gibt Aufschluß über die Entwicklung der Patientenzahlen, stratifiziert nach der Therapiezugehörigkeit. Bereits nach 8 Monaten Verweildauer hat sich die Zahl der unter Beobachtung stehenden Patienten von ursprünglich über 65 auf 10 in jeder Therapiegruppe reduziert. Ab dem 20. Monat ist kein Patient mehr unter Risiko, der mit der Standardchemotherapie behandelt wurde.

Um die Daten zu analysieren, haben wir wie auch Aydemir (1992) ein dynamisches binomiales Logit-Modell nach (3.14) verwendet:

$$\lambda(t|x_U) = \frac{\exp(\eta_{tU})}{1 + \exp(\eta_{tU})}, \quad \eta_{tU} = \tau_t + \mathbf{x}_{K1,U}\beta_{2,t} + \mathbf{x}_{K2,U}\beta_{3,t} + \mathbf{x}_{G,U}\beta_{4,t}$$

$$t = 1, \dots, \bar{T} = 34, U = 1, \dots, \bar{U} = 137. \quad (6.15)$$

τ_t ist dabei der Baseline-Parameter. Als Kovariable beziehen wir den Karnofsky-Index ein, dessen Einfluß sich in Kalbfleisch und Prentice (1980) als signifikant herausgestellt hat. Die Kodierung der Kovariablen Karnofsky-Index ist in Tabelle 6.12 zusammengefaßt:

Wert des Karnofsky-Index	Dummy x_{K1}	Dummy x_{K2}
10 - 30	1	0
40 - 60	0	1
70 - 90	0	0

Tabelle 6.12 Kodierung des Karnofsky-Index

Um zu erfahren, welchen Einfluß die Therapie hat, beziehen wir noch die Kovariable Therapiezugehörigkeit

$$x_G = \begin{cases} 1, & \text{Standardchemotherapie} \\ 0, & \text{Testchemotherapie} \end{cases}$$

in die Analyse ein. $\beta_{i,t}$, $i = 2,3,4$, sind die zugehörigen Effekte. Bis auf die Kovariable histologischer Tumortyp x_{Tu} stellten sich nach Kalbfleisch und Prentice (1980, Abschnitt 3.8.2) alle im Datensatz enthaltenen Kovariablen als nicht signifikant heraus.

Als Übergangsmodell für die Zustandsvektoren $\alpha_t = (\tau_t, \beta_{2,t}, \beta_{3,t}, \beta_{4,t})'$ mit $n = \dim(\alpha_t) = 4$ verwenden wir in diesem Teilabschnitt den RW1:

$$\begin{aligned} \alpha_t &= \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(\mathbf{0}_4, Q), \quad t = 1, \dots, \bar{T} = 34. \\ Q &\text{ diagonal, Verteilung von } \alpha_0 : \text{diffus.} \end{aligned} \quad (6.16)$$

Der Programmablauf der Posteriori-Modus-Schätzung mit dem Programm PME.ST ist in Tabelle 6.12 zusammengefaßt. In FS-Iter. steht wieder die insgesamt durchgeführte Anzahl von Fisher-Scoring-Schritten (durchschnittlich 6 FS-Schritte pro EM-Iteration).

Veteran's Administration Lungenkrebsdaten: Dynamisches binomiales Logit-Modell mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
$4.6 \cdot 10^{-5}$	$4.568 \cdot 10^{-5}$	1302	$1 \cdot 10^{-10}$	217	$2 \cdot 10^{-5}$
0.00115	0.0011478				
0.063	0.16				
0.024	0.08				

Tabelle 6.13: Hyperparameterschätzung und Programmablauf von PME.ST.

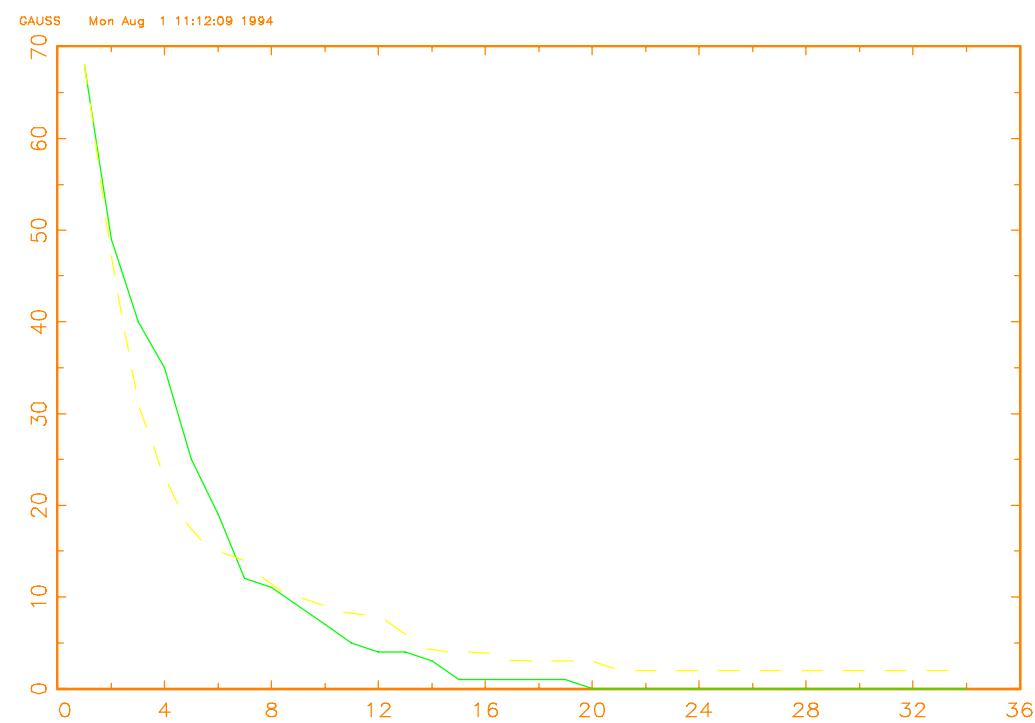


Abbildung 6.24: Anzahl der Patienten, die unter Risiko stehen (nicht gestorben oder zensiert): — Standardchemotherapie, - - - Testchemotherapie.
Abszisse: Zeit t in Monaten.

Abbildung 6.25 enthält den Baseline-Parameter, der als zeitkonstant geschätzt wurde. Die Karnofsky-Effekte sind in Abbildung 6.26 zusammengefaßt. Die Karnofsky-Werte 0-30 haben einen zeitkonstanten, signifikant positiven Einfluß auf die Hazardrate. Das bedeutet, daß Patienten mit schlechtem Gesundheitszustand ein höheres Sterberisiko haben im Vergleich zu Patienten mit Karnofsky-Werten von 70-90. Für Karnofsky-Werte von 40-60 ergibt sich bis zum 4. Monat ein signifikant positiver Einfluß auf die Hazardrate im Vergleich zur Karnofsky-Gruppe mit Werten 70-90. Ab dem 4. Monat ist die Null-niveaulinie stets innerhalb des punktwisen Konfidenzbandes, so daß kein signifikanter Einfluß mehr vorliegt.

Der geschätzte Therapieeffekt ist in Abbildung 6.27 angegeben. Zu Beginn der Untersuchung hat er einen Wert von $\beta_{4,1|34} \approx -0.2$. Dieser Wert stimmt sehr gut mit dem Resultat in Kalbfleisch und Prentice (1980, Abschnitt 3.8.2) überein, das im Rahmen eines Weibull Regressionsmodells mit zeitkonstanten Effekten erzielt wurde. Dort war lediglich die Referenzkategorie für x_G anders gewählt: $x_G = 0$ für die Standardchemotherapie und $x_G = 1$ für die Testchemotherapie. In unserem Modell (6.15), (6.16) lassen wir im Gegensatz zu der Analyse in Kalbfleisch und Prentice (1980) zu, daß der Therapieeffekt über die Zeit variiert. Mit zunehmender Verweildauer steigt der Effekt in Abbildung 6.27

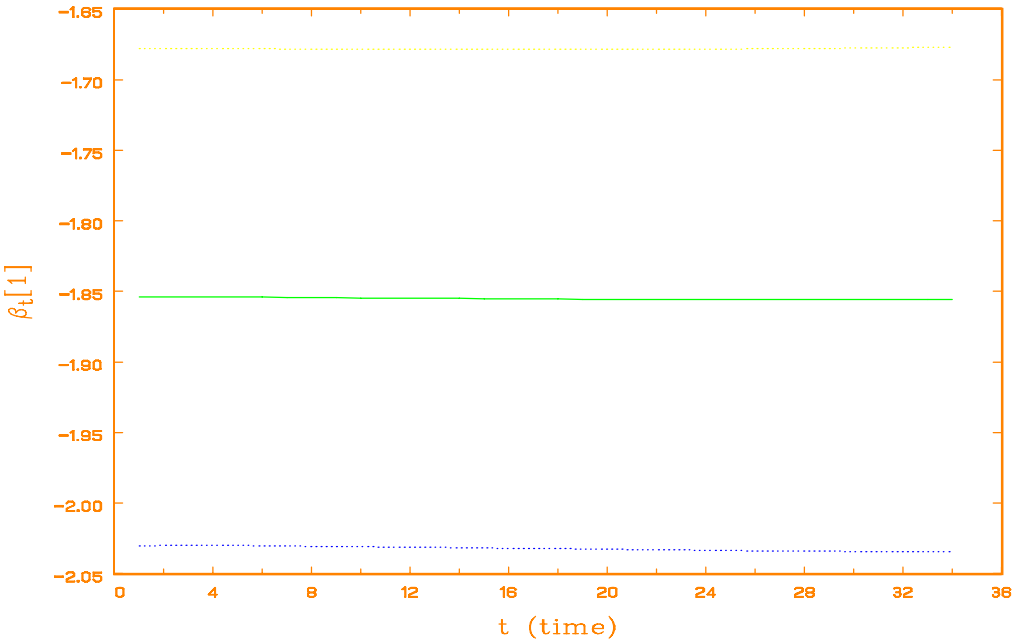


Abbildung 6.25: Geschätzter Baseline-Parameter $\tau_t = \beta_{1,t}$ mit Konfidenzband, dynamisches binomiales Logit-Modell mit RW1. Zeit t in Monaten.

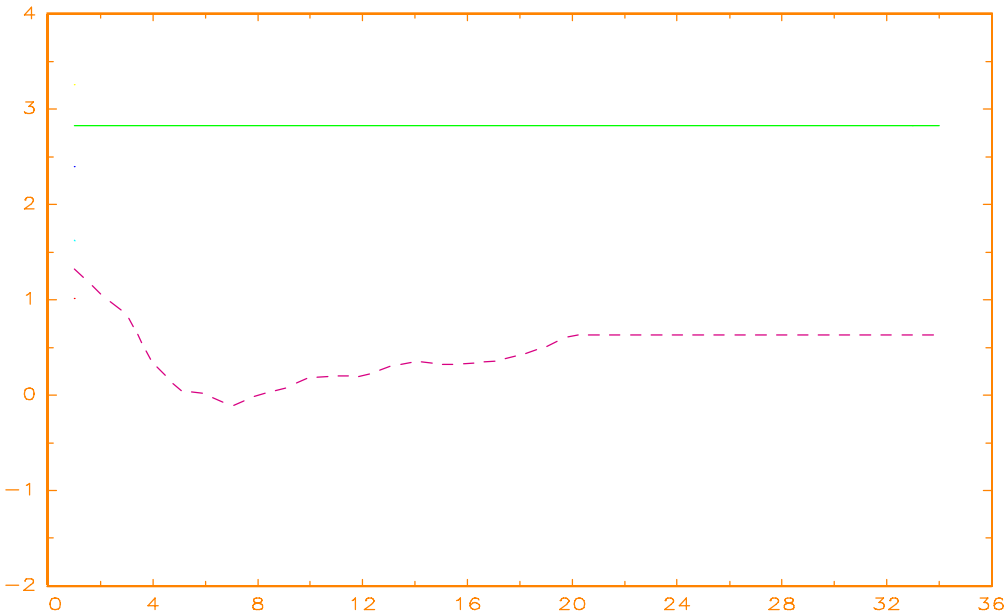


Abbildung 6.26: Geschätzte Karnofsky-Effekte mit Konfidenzbänder: — $\beta_{2,t}$,
--- $\beta_{3,t}$ dynamisches binomiales Logit-Modell mit RW1. Zeit t in
Monaten.

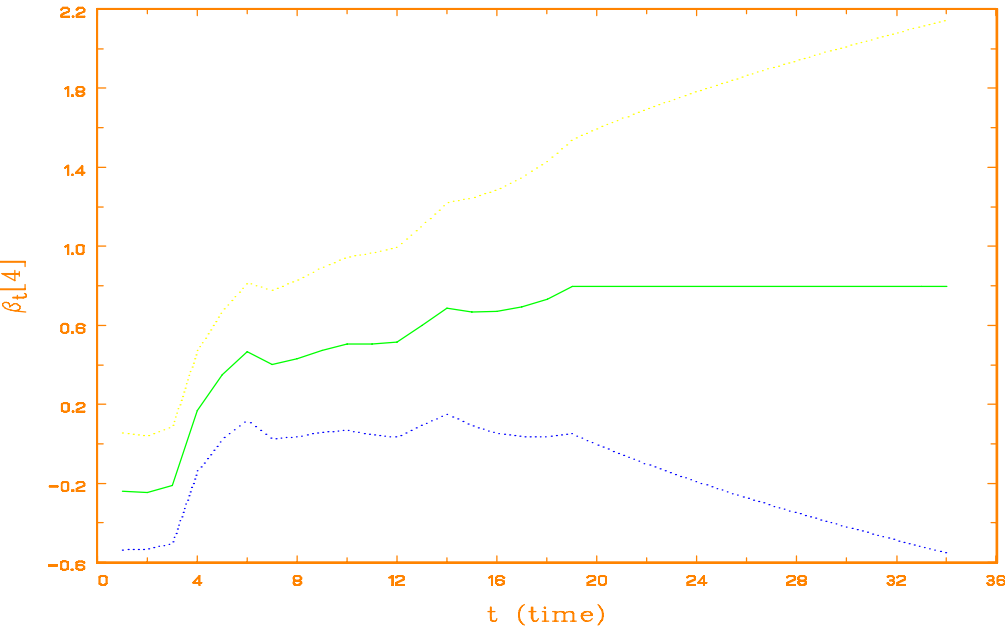


Abbildung 6.27: Geschätzter Therapieeffekt $\beta_{4,t}$ mit Konfidenzband, dynamisches
binomiales Logit-Modell mit RW1. Zeit t in Monaten.

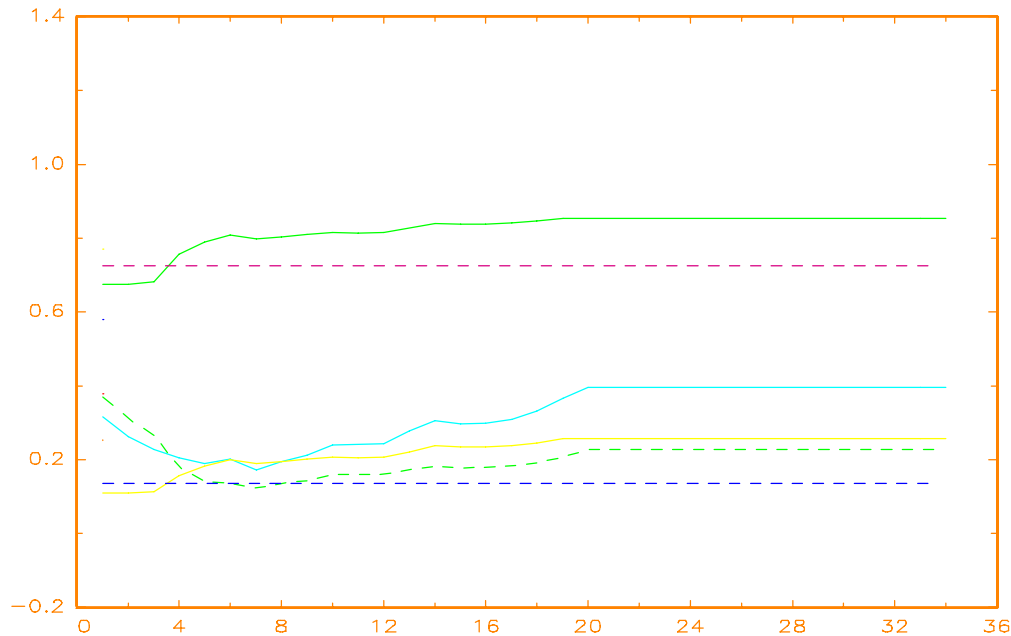


Abbildung 6.28: Geschätzte Hazardraten: — Standardchemotherapie, - - - Testchemotherapie. Obere Kurven: Karnofsky 10-30, mittlere Kurven: Karnofsky 40-60, untere Kurven: Karnofsky 70-90. Dynamisches binomiales Logit-Modell mit RW1. Abszisse: Zeit t in Monaten.

weiter an, schneidet zwischen dem 3. und 4. Monat die Nullniveauinie und bleibt bis zum Ende des Beobachtungszeitraumes positiv. Das bedeutet, die Standardchemotherapie hat zu Beginn einen negativen und ab dem 4. Monat einen positiven Einfluß auf die Hazardrate im Vergleich zur Testchemotherapie. Der Einfluß dieser Kovariable Therapiezugehörigkeit ist jedoch nicht beziehungsweise nicht deutlich signifikant. Die Nullniveauinie liegt entweder innerhalb des punktwisen Konfidenzbandes oder für die Monate 5 bis 20 knapp unterhalb des unteren Konfidenzbandes. Ab dem 20. Monat ist der geschätzte Effekt konstant. Dies hängt damit zusammen, daß ab diesem Monat kein Patient aus der Standardchemotherapiegruppe, wie wir in Abbildung 6.24 gesehen haben, unter Beobachtung steht. Die geschätzten Hazardraten sind in Abbildung 6.28 widergegeben. Konfidenzbänder wurden aus Gründen der Übersichtlichkeit nur für die Hazardrate (Standardchemotherapie und Karnofsky-Wert 10-30) und die Hazardrate (Standardchemotherapie und Karnofsky-Wert 40-60) angegeben. Die Schätzungen für die Hazardraten entsprechen den Ergebnissen aus der Effektschätzung. Für Karnofsky-Werte 10-30 sind die Hazardraten signifikant am

höchsten, für Karnofsky-Werte 40-60 am zweithöchsten, allerdings kaum über den Hazardraten für Karnofsky-Werte 70-90. Die Hazardraten für die Testchemotherapiegruppe liegen alle, abgesehen von den ersten 4 Monaten, unter den Kurven für die Standardchemotherapie.

Als Ergebnis bleibt festzuhalten, daß die Standardchemotherapie zu Beginn die bessere Therapie ist. Ab dem 4. Monat ändert sich die Situation, und die Testchemotherapie weist niedrigere Hazardratenverläufe im Vergleich zur Standardchemotherapie auf. Der Einfluß der Therapie ist jedoch nicht signifikant.

6.4 Gliomdaten

Zur Analyse dieses Datensatzes wurden die Verweildauern, die in Tagen vorliegen, zu Monaten gruppiert. Wie in Aydemir (1992) verwenden wir das dynamische gruppierte Cox-Modell

$$\lambda(t|x_U) = 1 - \exp\{-\exp(\eta_{tU})\},$$

$$\begin{aligned} \eta_{tU} = & \tau_t + \mathbf{x}_{G,U}\beta_{2,t} + \mathbf{x}_{M1,U}\beta_{3,t} + \mathbf{x}_{M2,U}\beta_{4,t} + \mathbf{x}_{A,U}\beta_{5,t} \\ & + \mathbf{x}_{K1,U}\beta_{6,t} + \mathbf{x}_{K2,U}\beta_{7,t} + \mathbf{x}_{R,U}\beta_{8,t} + \mathbf{x}_{E,U}\beta_{9,t} \end{aligned}$$

$$t = 1, \dots, \bar{T} = 73, U = 1, \dots, \bar{U} = 467. \quad (6.17)$$

Tabelle 6.14 zeigt, wie die Referenzkategorien festgelegt sind:

Kovariablen	\mathbf{x}_G	\mathbf{x}_M	\mathbf{x}_A	\mathbf{x}_K	\mathbf{x}_R	\mathbf{x}_E
Referenzkategorie	BCNU+ VM26	Grad 3-4	> 50 Jahre	≥ 90%	reseziert	Epi.-Anfall: nein

Tabelle 6.14 Referenzkategorien für die Kovariablen im Gliomdatensatz.

Als Übergangsmodell für die Zustandsvektoren $\alpha_t = (\tau_t, \beta_{2,t}, \beta_{3,t}, \beta_{4,t}, \beta_{5,t}, \beta_{6,t}, \beta_{7,t}, \beta_{8,t}, \beta_{9,t})'$ mit $n = \dim(\alpha_t) = 9$ verwenden wir in diesem Teilabschnitt den RW1:

$$\alpha_t = \alpha_{t-1} + \xi_t, \quad \xi_t \sim N(\mathbf{0}_g, Q), \quad t = 1, \dots, \bar{T} = 73.$$
$$Q \text{ diagonal, Verteilung von } \alpha_0 : \text{diffus.} \tag{6.18}$$

Die Tabelle 6.15 gibt über den algorithmischen Ablauf der Schätzung Auskunft:

Gliomdaten: Dynamisches gruppiertes Cox-Modell mit RW1					
$Q^{(0)}$	$\text{diag}(\hat{Q})$	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
$0.01 \cdot I_9$	0.02825	850	$1 \cdot 10^{-10}$	85	$5 \cdot 10^{-4}$
	0.00362				
	0.00278				
	0.00715				
	0.00464				
	0.00753				
	0.00276				
	0.00429				
	0.00879				

Tabelle 6.13: Hyperparameterschätzung und Programmablauf von PME.ST.

Der geschätzte Baseline-Parameter und die geschätzten Effekte sind in den Abbildungen 6.29 bis 6.35 dargestellt. Der Einfluß der BCNU-Therapie auf die Hazardrate im Vergleich zur kombinierten BCNU+VM26-Therapie, der durch den Therapieeffekt in Abbildung 6.30 angegeben wird, steigt in den ersten 20 Monaten an und verläuft in der nachfolgenden Beobachtungszeit nahezu konstant. In den Monaten 12 bis 25 liegt das untere Konfidenzband knapp über dem Nullniveau. Ansonsten befindet sich die Nullniveaulinie stets im Konfidenzbereich. In der Abbildung 6.31 sind die Effekte der Malignität geplottet. Der Malignitätsgrad 3 übt einen signifikant negativen, nahezu zeitkonstanten Einfluß auf die Hazardrate aus, während der hohe Malignitätsgrad 4 in den ersten 30 Monaten einen signifikant positiven Effekt im Vergleich zu Malignitätsgrad 3-4 besitzt. Gegen Ende des Untersuchungszeitraumes verringert sich dieser Effekt und das Konfidenzband wird sehr breit. Aus der Abbildung 6.32 wird deutlich, daß sich „junges Alter“ (≤ 50 Jahre) im

Vergleich zur Referenzkategorie signifikant negativ auf die Hazardrate und damit positiv auf die Überlebensrate auswirkt. In der Abbildung 6.33 sind die Karnofsky-Effekte dargestellt. Der Effekt eines niedrigen Karnofsky-Wertes im Vergleich zur Referenzgruppe ($\geq 90\%$) ist zu Beginn signifikant und stark positiv. Im weiteren Verlauf schwächt er sich ab bis zum 20. Monat. Von da an bleibt er nahezu unverändert positiv, aber nicht mehr signifikant. Die mittleren Karnofsky-Werte besitzen im Vergleich zur Referenzkategorie einen leicht positiven, aber nicht signifikanten Einfluß auf die Hazardrate. Der Effekt der Biopsie in Abbildung 6.34 ist signifikant positiv und zeigt wenig Variation in der Zeit. Das bedeutet, daß Patienten, bei denen aufgrund ihres fortgeschrittenen Krankheitszustandes eine Operation des Tumors nicht möglich war, eine signifikant schlechtere Überlebenschance besitzen als Patienten, bei denen der Tumor teilweise operativ entfernt wurde. Aus der Abbildung 6.35 geht hervor, daß epileptische Anfälle einen signifikant negativen Einfluß auf die Hazardrate und damit einen positiven auf das Überleben haben.

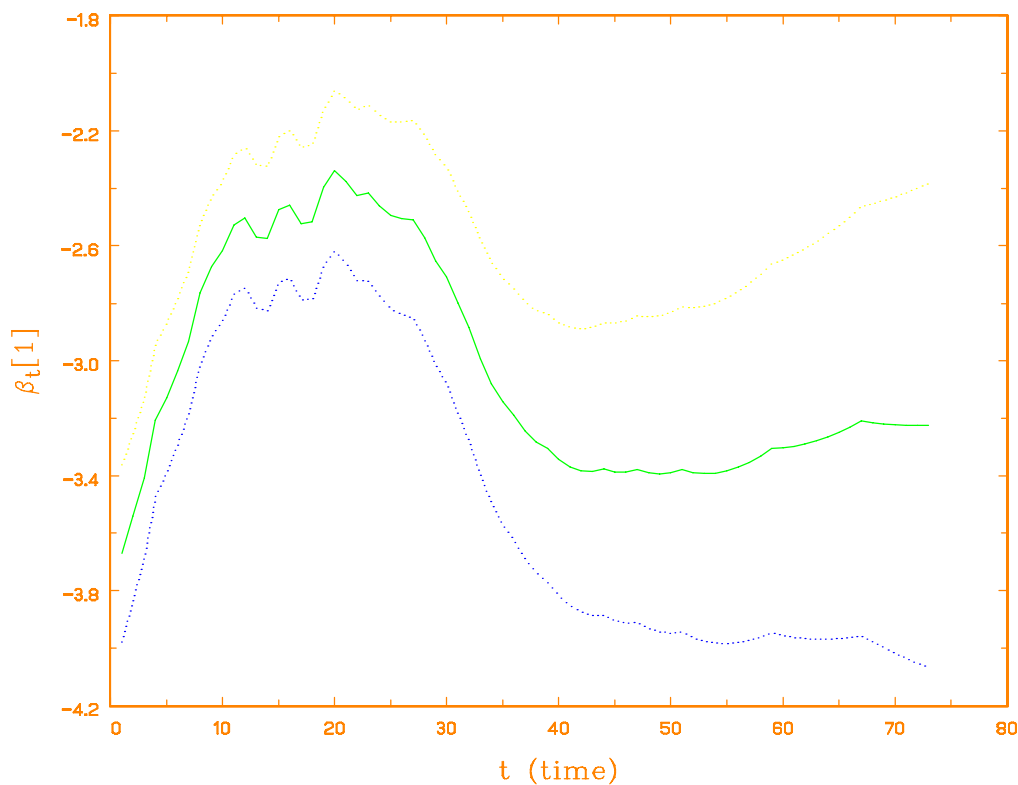


Abbildung 6.29: Geschätzter Baseline-Parameter $\tau_t = \beta_{1,t}$ mit Konfidenzband, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten.

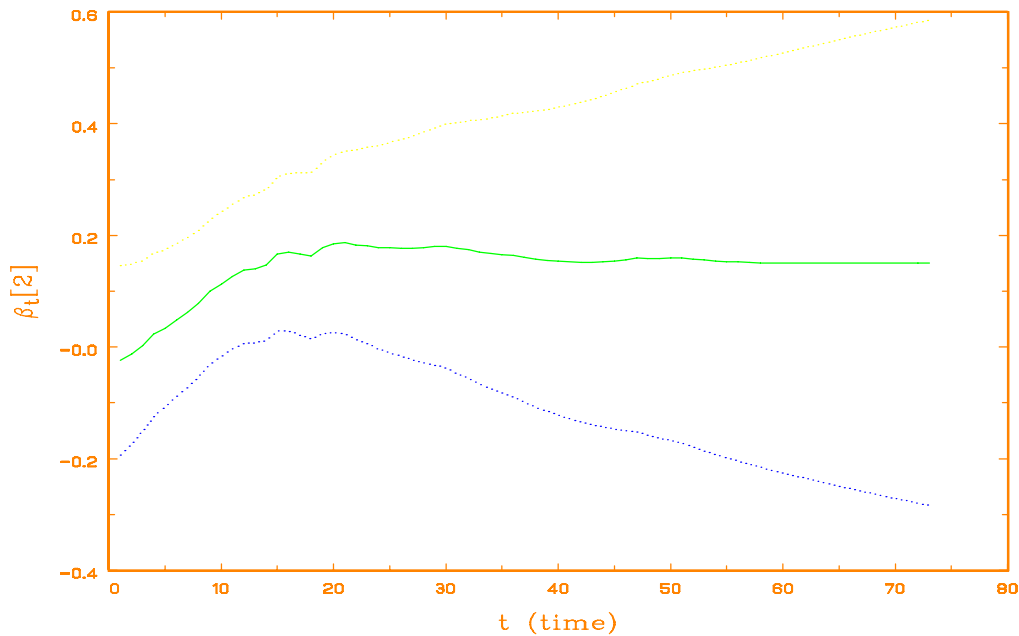


Abbildung 6.30: Geschätzter Therapieeffekt mit Konfidenzband, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten.

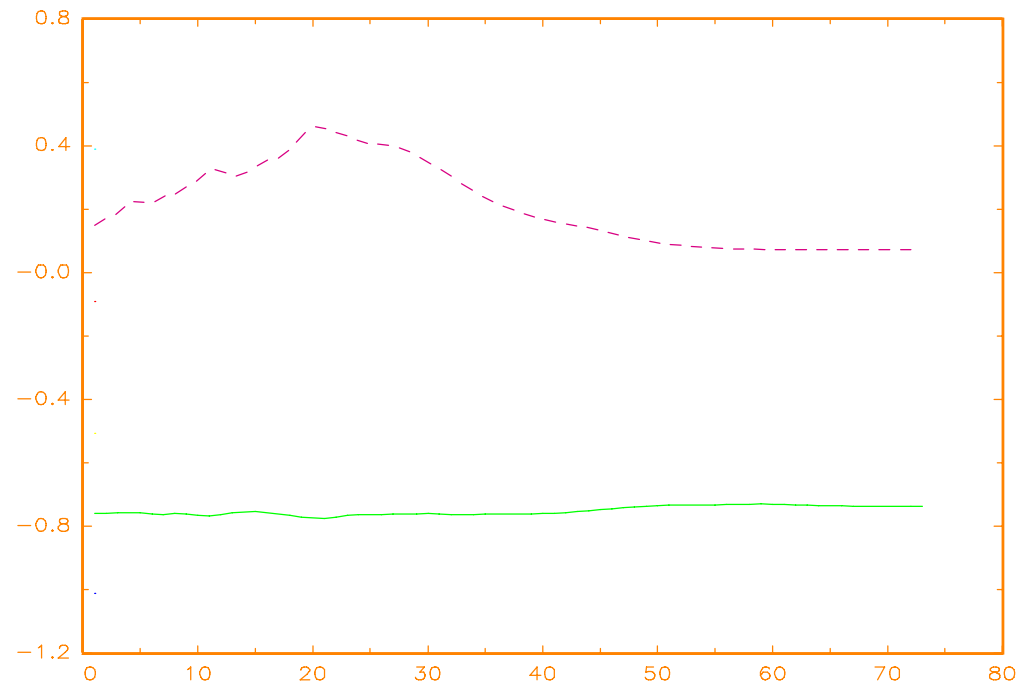


Abbildung 6.31: Geschätzte Effekte der Malignität mit Konfidenzband: — Grad 3, — — — Grad 4, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten.

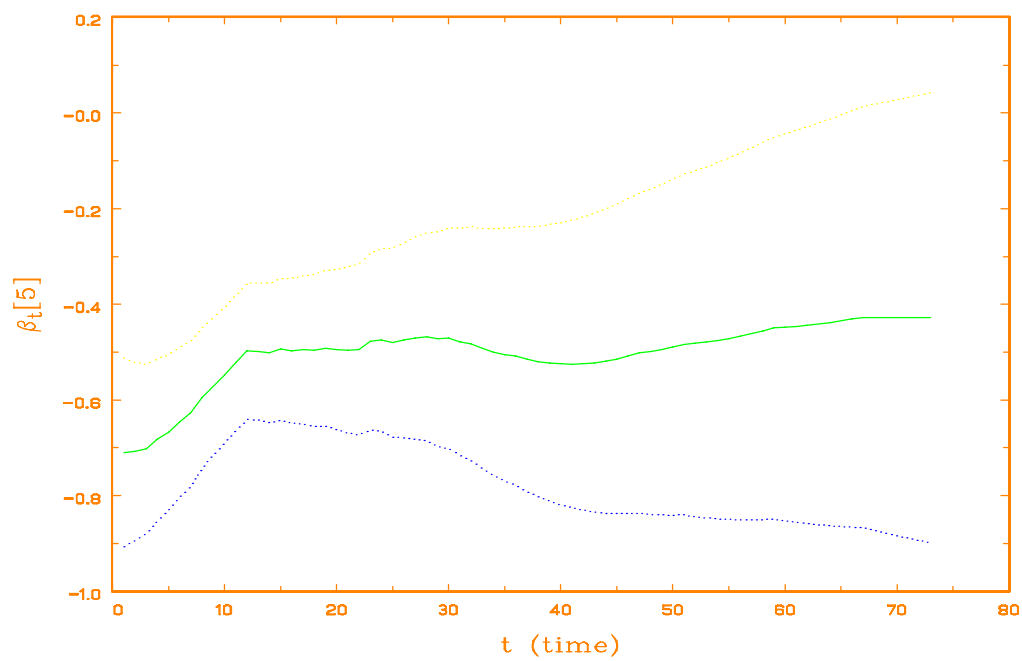


Abbildung 6.32: Geschätzter Alterseffekt mit Konfidenzband, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten.

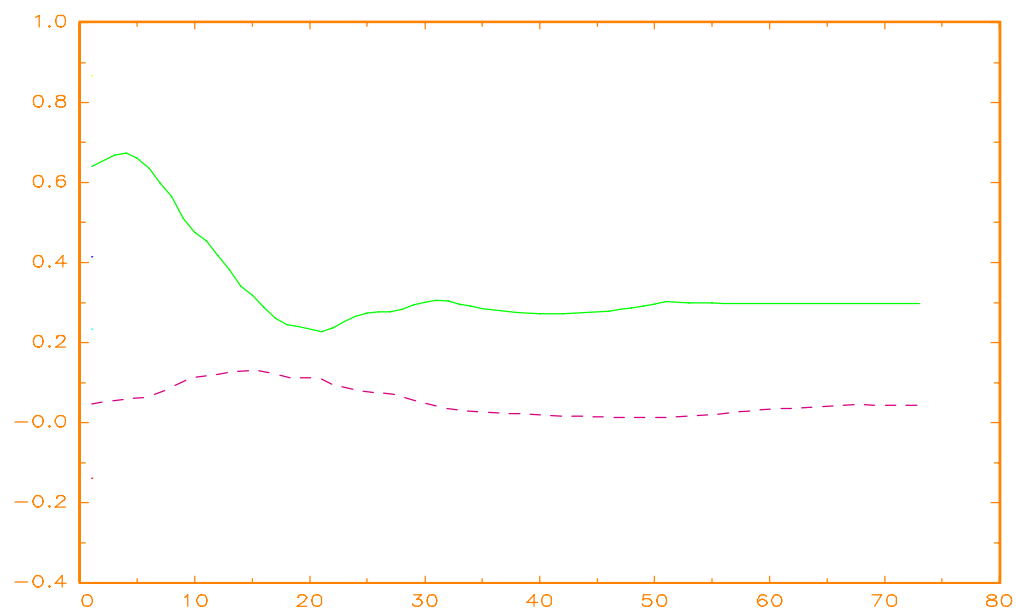


Abbildung 6.33: Geschätzte Effekte des Karnofsky-Index mit Konfidenzband: — Karnofsky-Index $\leq 60\%$, - - - - Karnofsky-Index 70-80%, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten.

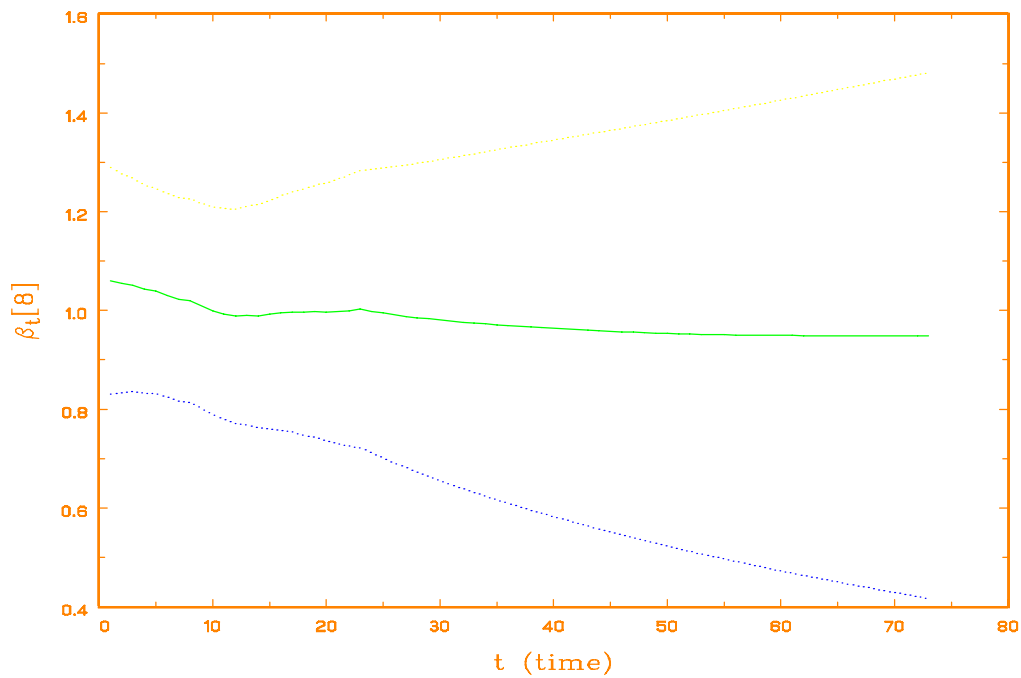


Abbildung 6.34: Geschätzter Effekt der Resektion mit Konfidenzband, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten.

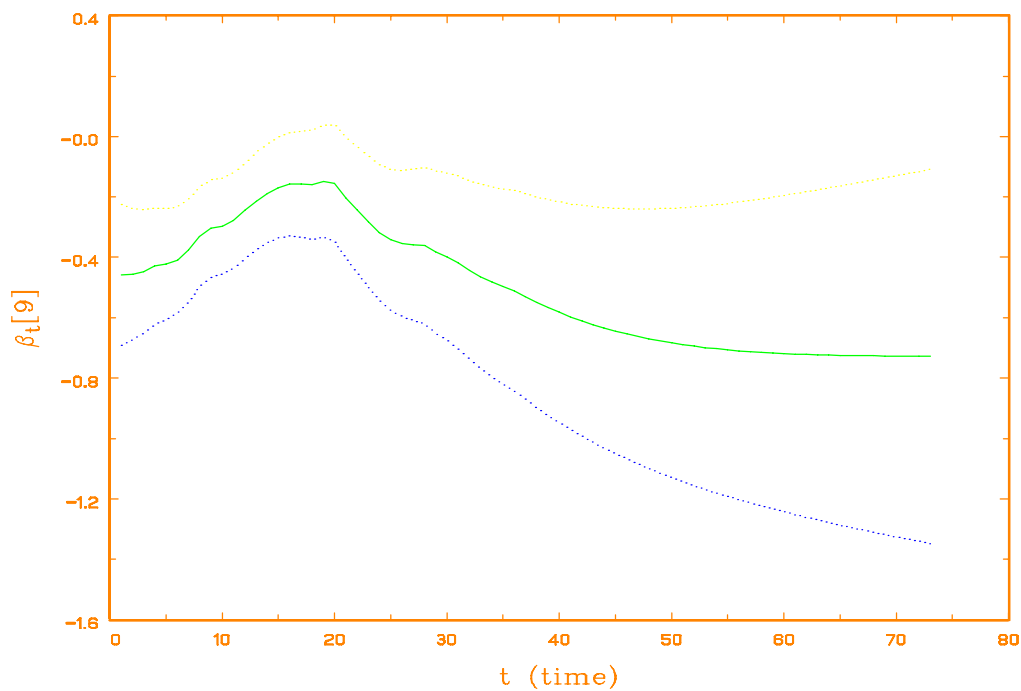


Abbildung 6.35: Geschätzter Effekt der epileptischen Anfälle mit Konfidenzband, dynamisches gruppiertes Cox-Modell mit RW1. Zeit t in Monaten.

In der Abbildung 6.36 sind die Hazardraten von zwei Personen exemplarisch dargestellt. Die Person Nummer 1 hat die BCNU-Therapie erhalten, besitzt den Malignitätsgrad 3 und einen Karnofsky-Index $\leq 60\%$, ist über 50 Jahre alt, der Tumor wurde teilweise operativ entfernt und es sind epileptische Anfälle aufgetreten. Bei der Person Nummer 2 wurde die kombinierte BCNU-VM26-Therapie angewendet. Ferner besitzt sie den Malignitätsgrad 3 und einen Karnofsky-Index 70-80%, ist ≤ 50 Jahre alt, wurde biopsiert und es sind keine epileptischen Anfälle aufgetreten.

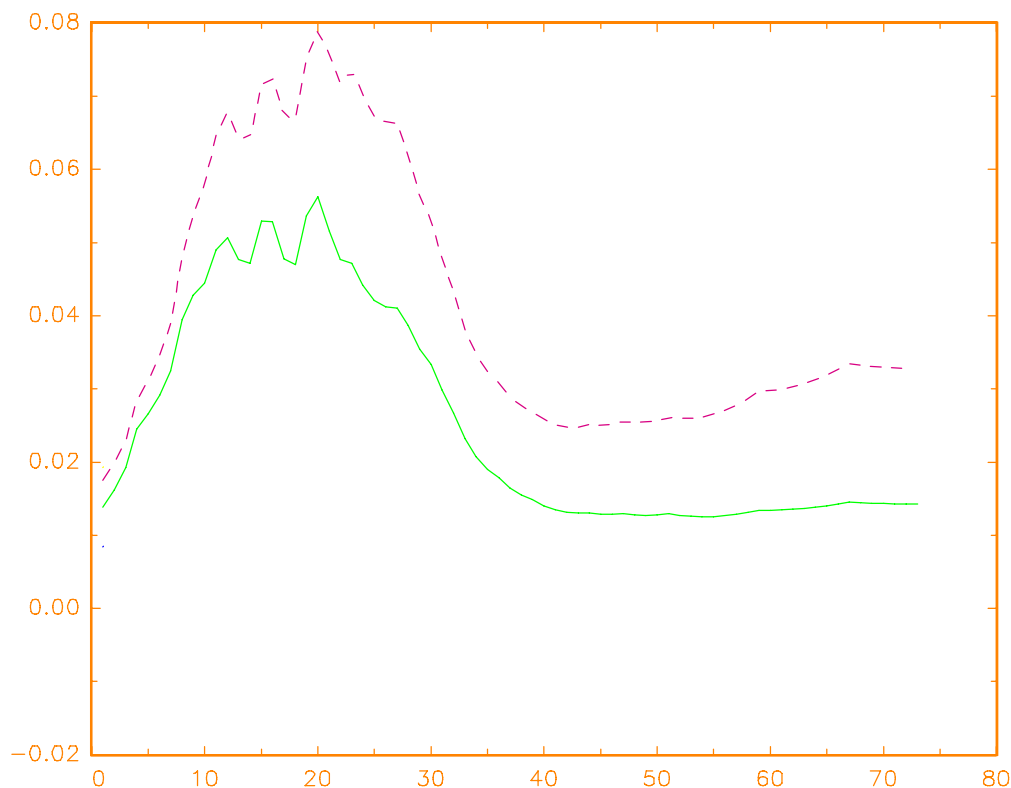


Abbildung 6.36: Geschätzte Hazardraten mit Konfidenzband: — Person Nummer 1, — — — Person Nummer 2, dynamisches gruppiertes Cox-Modell mit RW1. Abszisse: Zeit t in Monaten.

Die geschätzte Hazardrate für die Person Nummer 2 liegt, abgesehen von den ersten 12 Monaten, signifikant über der von Person Nummer 1. Dies ist wohl auf den sehr starken Alterseffekt zurückzuführen. Weil sich die Personen in den meisten anderen Kovariablen,

außer dem mit sehr starken Einfluß versehenen Malignitätsgrad 3, ebenfalls unterscheiden, kam es insgesamt zu einer Überlagerung des Einflusses der Effekte. Das wiederum mag eine Erklärung dafür sein, daß die Hazardrate von Person Nummer 2 in den ersten 12 Monaten nicht signifikant über der von Person Nummer 1 liegt, obwohl der Alterseffekt in Abbildung 6.32 selbst signifikant ist.

Für eine Subgruppenanalyse haben wir den Datensatz in 2 Gruppen aufgeteilt:

1. Gruppe: Alter ≤ 50 Jahre mit 180 Patienten und
2. Gruppe: Alter > 50 Jahre mit den restlichen 287 Patienten.

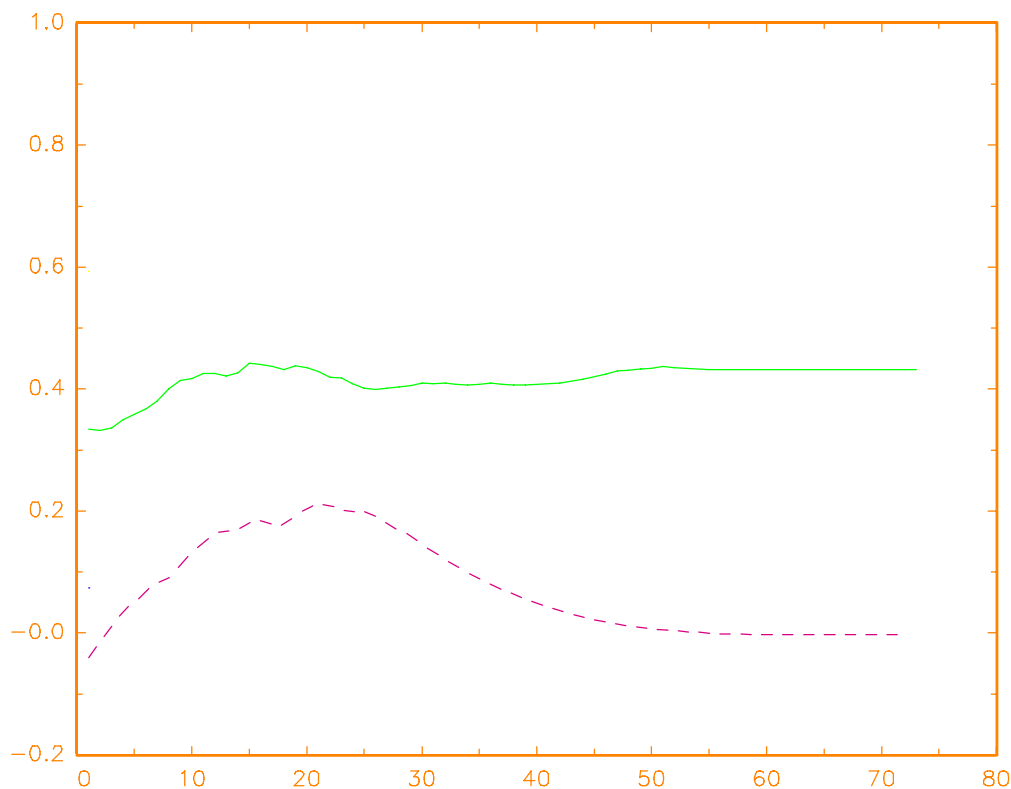


Abbildung 6.37: Geschätzte Therapieeffekte in den Altersgruppen: — Alter ≤ 50 Jahre, — — — Alter > 50 Jahre. Dynamisches gruppiertes Cox-Modell mit RW1. Abszisse: Zeit t in Monaten.

Für die Analyse haben wir wieder das dynamische gruppierte Cox-Modell verwendet, allerdings nur mit der Kovariablen Therapiezugehörigkeit. Als Übergangsmodell für die

Parameter wurde der RW1 benutzt. Das Ziel war es zu erkunden, wie sich der Therapieeffekt in den einzelnen Subgruppen verhält. Die Abbildung 6.37 zeigt die geschätzten Therapieeffekte in den beiden Gruppen. Der Therapieeffekt in der Altersgruppe ≤ 50 Jahre liegt meist signifikant über dem Nullniveau und dem Effekt für die zweite Altersgruppe. Außerdem ist er auch höher als der im Rahmen der Gesamtanalyse erhaltene Therapieeffekt in der Abbildung 6.30. Das bedeutet, bei den jüngeren Patienten hat die BCNU-Therapie einen noch stärkeren, positiven Einfluß auf die Hazardrate und damit das Sterberisiko als bei den älteren. Deshalb sollte nach dieser Untersuchung gerade für die Patienten in der Altersgruppe 1 die kombinierte BCNU+VM26-Therapie angewendet werden.

6.5 SOEP Arbeitslosendaten

Im folgenden analysieren wir die Arbeitslosendaten aus dem SOEP zunächst getrennt nach dem Ein-Episoden-Ein-Zustands-Fall, dem Mehr-Zustands-Fall und dem Mehr-Episoden-Fall. Ein Vergleich der Analysemethoden anhand der Berechnungen und Schätzergebnisse zeigt deren Vor- und Nachteile

6.5.1 Ein-Episoden-Ein-Zustands-Analyse

Im Ein-Episoden-Ein-Zustands-Fall interessieren wir uns für die Verweildauer im Zustand Arbeitslosigkeit und für den Einfluß von bestimmten Kovariablen. In unserer Untersuchung haben wir die folgenden personenspezifischen Kovariablen berücksichtigt:

- Nationalität: West-Deutscher ($N1 = 1$, 0 sonst)
 Ost-Deutscher ($N2 = 1$, 0 sonst)
 Ausländer, West ($N1 = N2 = 0$) : Referenzgruppe
- Geschlecht: Mann ($S = 1$)
 Frau ($S = 0$) : Referenzgruppe
- Alter: Alter ≤ 30 Jahre ($A1 = 1$, 0 sonst)
 $41 \leq \text{Alter} \leq 50$ Jahre ($A2 = 1$, 0 sonst)
 Alter ≥ 51 Jahre ($A3 = 1$, 0 sonst)
 Referenzgruppe: $31 \leq \text{Alter} \leq 40$ Jahre

Für die Analyse verwenden wir ein dynamisches binomiales Logit-Modell

$$\lambda(t|\mathbf{x}_U) = \frac{\exp(\tau_t + \mathbf{x}'_U \beta_t)}{1 + \exp(\tau_t + \mathbf{x}'_U \beta_t)}, \quad t = 1, \dots, \bar{T} = 36, U = 1, \dots, \bar{U} = 2911.$$

(6.19)

Der Dummy-Vektor $\mathbf{x}_U = (\mathbf{N1}_U, \mathbf{N2}_U, \mathbf{S}_U, \mathbf{A1}_U, \mathbf{A2}_U, \mathbf{A3}_U)'$ enthält in Dummy-Kodierung die Kovariablen Nationalität, Geschlecht und Alter. Der Baseline-Parameter τ_t und der Vektor der Effekte $\beta_t = (\beta_{2,t}, \beta_{3,t}, \beta_{4,t}, \beta_{5,t}, \beta_{6,t}, \beta_{7,t})'$ werden im Übergangsmodell mit einem RW1 modelliert.

Die Tabelle 6.14 faßt den Programmablauf zusammen.

Arbeitslosendaten, Ein-Episoden-Ein-Zustands-Fall: Dynamisches binomiales Logit-Modell mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.0059	0.2638	567	$1 \cdot 10^{-10}$	81	$2 \cdot 10^{-4}$
0.0285	0.0059				
0.0010	0.0095				
0.0099	0.0114				
0.0039	0.0035				
0.0084	0.0038				
0.0096	0.0616				

Tabelle 6.14: Hyperparameterschätzung und Programmablauf von PME.ST.

Die Abbildungen 6.38 bis 6.43 enthalten die geschätzten Effekte. Der Effekt, West-Deutscher zu sein, wirkt sich in den ersten zwei Jahren im Vergleich zu einem Ausländer (West) positiv auf die Hazardrate und damit auf den Abgang aus der Arbeitslosigkeit aus. Der zu Beginn stark positive Effekt in Abbildung 6.38 schwächt sich allerdings mit der Zeit ab. Auch der Effekt Ost-Deutscher im Vergleich zu Ausländer (West) in Abbildung 6.39 ist zu Beginn der Arbeitslosigkeit signifikant positiv. Bei längerer Arbeitslosigkeit schwächt sich auch hier der Effekt auf die Hazardrate ab und schneidet nach 12 Monaten die

Nullniveaulinie. Nach einem Jahr Arbeitslosigkeit ist der Effekt negativ, die Nullniveaulinie liegt jedoch im Konfidenzbereich. Auch der Geschlechtseffekt in Abbildung 6.40 ist zu Beginn signifikant positiv, schwächt sich mit zunehmender Arbeitslosigkeitsdauer ab und schneidet bei 9 Monaten die Nullniveaulinie. Danach ist das obere Konfidenzband knapp unterhalb dem Nullniveau. Gegen Ende ist ein leichter Anstieg des Effekts zu verzeichnen, allerdings bei einem sehr breiten Konfidenzband. Das bedeutet, daß Männer nach kurzer Arbeitslosigkeitsdauer (weniger als 9 Monate) eine höhere Abgangsrate aus der Arbeitslosigkeit besitzen als Frauen. Bei längerer Arbeitslosigkeitsdauer ändert sich die Situation. Der Alterseffekt der Kovariablen A1 in Abbildung 6.41 ist zu Beginn signifikant positiv. Jüngere Arbeitslose haben, wie zu erwarten, bessere Chancen auf dem Arbeitsmarkt als ältere. Der Effekt schwächt sich mit zunehmender Arbeitslosigkeitsdauer ab, bleibt aber positiv. Vom 18. Monat an liegt die Nullniveaulinie im Konfidenzbereich. Der Alterseffekt der Kovariablen A2 in der Abbildung 6.42 ist signifikant negativ. Er zeigt eine leichte zeitliche Variation. Überraschend ist der Verlauf des Alterseffektes der Kovariablen A3 in Abbildung 6.43. Zu Beginn ist der Effekt signifikant negativ. Im weiteren Verlauf der Arbeitslosigkeit steigt er tendenziell an. Ab dem 20. Monat der Arbeitslosigkeit liegt das Nullniveau im Konfidenzbereich. Das bedeutet, daß die über 50-jährigen bei kürzerer Arbeitslosigkeitsdauer im Vergleich zur Referenzgruppe der 31 bis 40-jährigen signifikant schlechtere Chancen auf dem Arbeitsmarkt haben. Nach zwei Jahren Arbeitslosigkeitsdauer und länger ist der Effekt nicht mehr signifikant und zeitweise sogar leicht positiv. Eine Erklärung für diesen Sachverhalt erhalten wir mit einer Mehr-Zustands-Analyse in Teilabschnitt 6.5.2.

Eine Besonderheit ist bei der Schätzung des Baseline-Parameters in Abbildung 6.44 festzustellen. Jeweils in den Zeitintervallen 12, 24, 36, also stets nach einem Jahr Arbeitslosigkeit, treten Spitzen auf. Eine Erklärung liefert die Abbildung 6.45. Sie zeigt für jeden Verweildauermonat in der Arbeitslosigkeit die Anzahl der Untersuchungseinheiten, die zensiert aus der Beobachtung ausscheiden. Auch hier sind die Spitzen nach 12, 24 und 36 Monaten zu sehen. Entsprechende Plots für die anderen Abgangsarten aus der Arbeitslosigkeit (1-3 und 5-9, siehe dazu Teilabschnitt 1.1.5) zeigen keine solchen Peaks. Damit liefert das retrospektive Erhebungsdesign des SOEP eine mögliche Erklärung für die Spitzen.

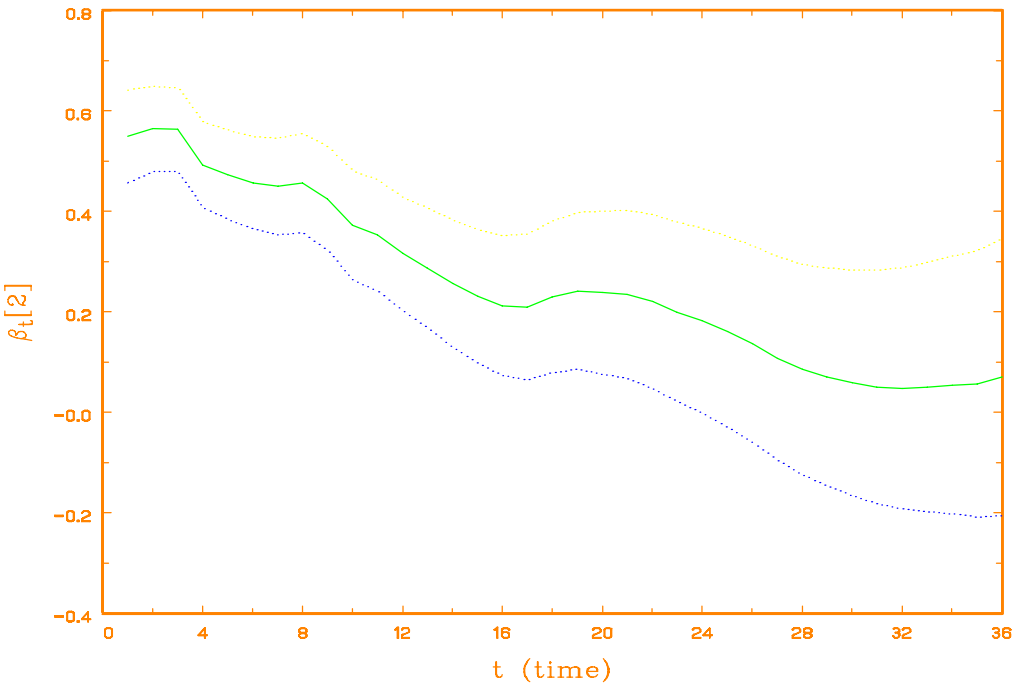


Abbildung 6.38 Geschätzter Nationalitätseffekt zur Kovariablen N1 mit Konfidenzband.

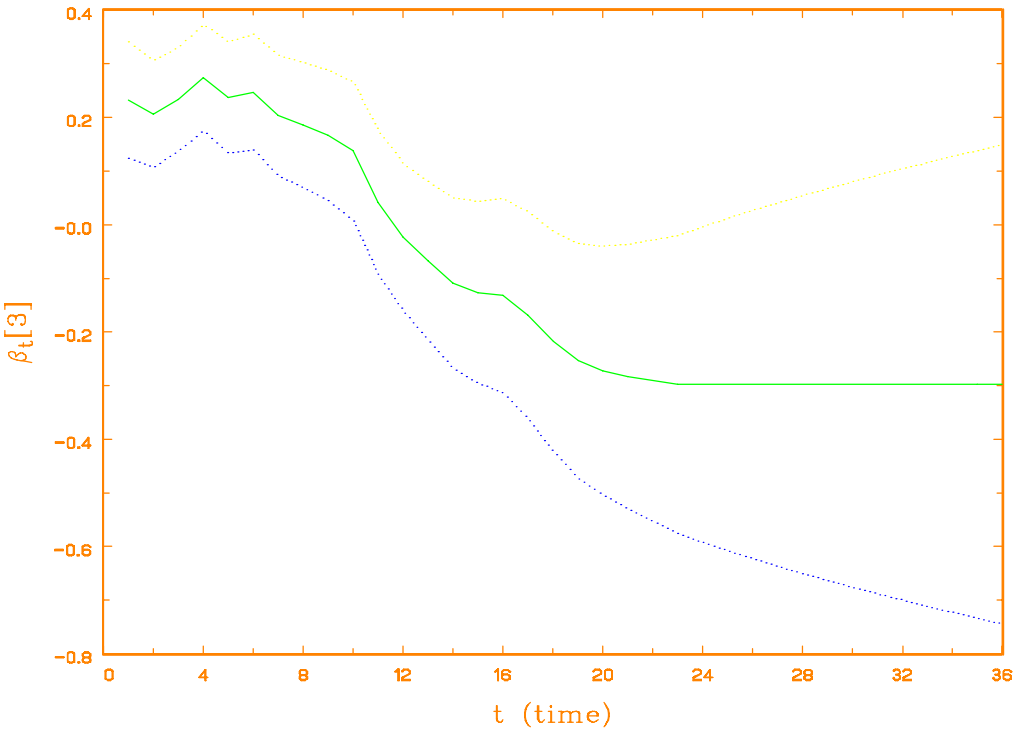


Abbildung 6.39 Geschätzter Nationalitätseffekt zur Kovariablen N2 mit Konfidenzband.

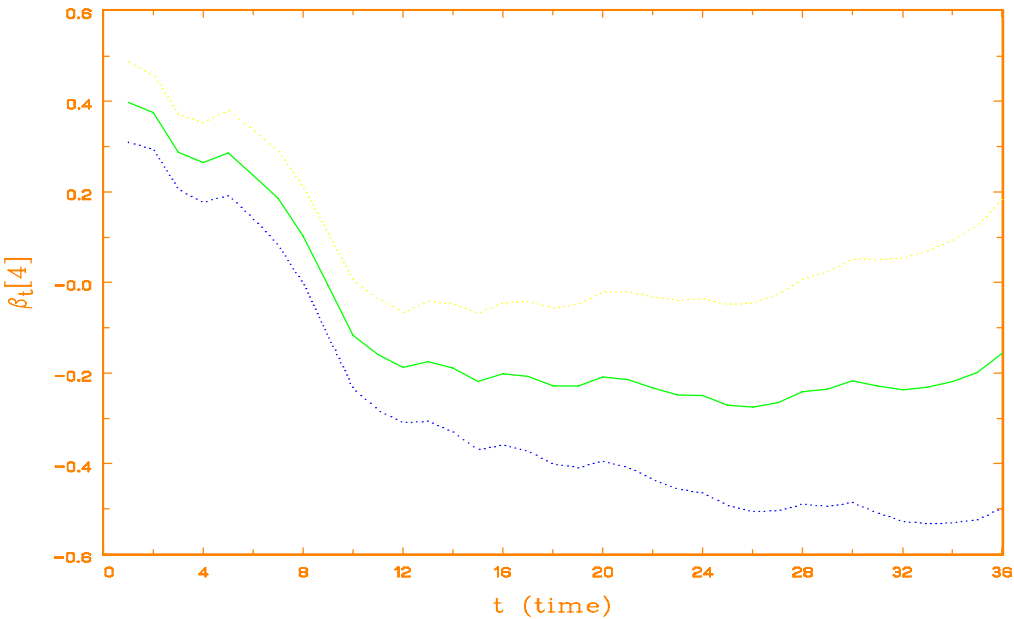


Abbildung 6.40 Geschätzter Geschlechtseffekt zur Kovariablen S mit Konfidenzband.

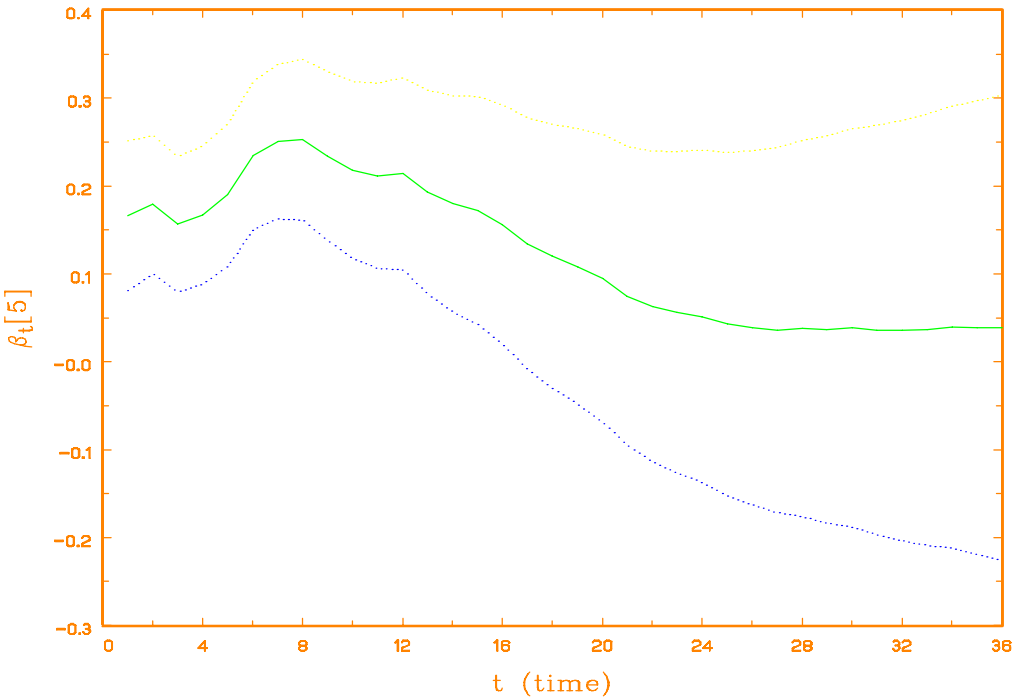


Abbildung 6.41 Geschätzter Altersseffekt zur Kovariablen A1 mit Konfidenzband.

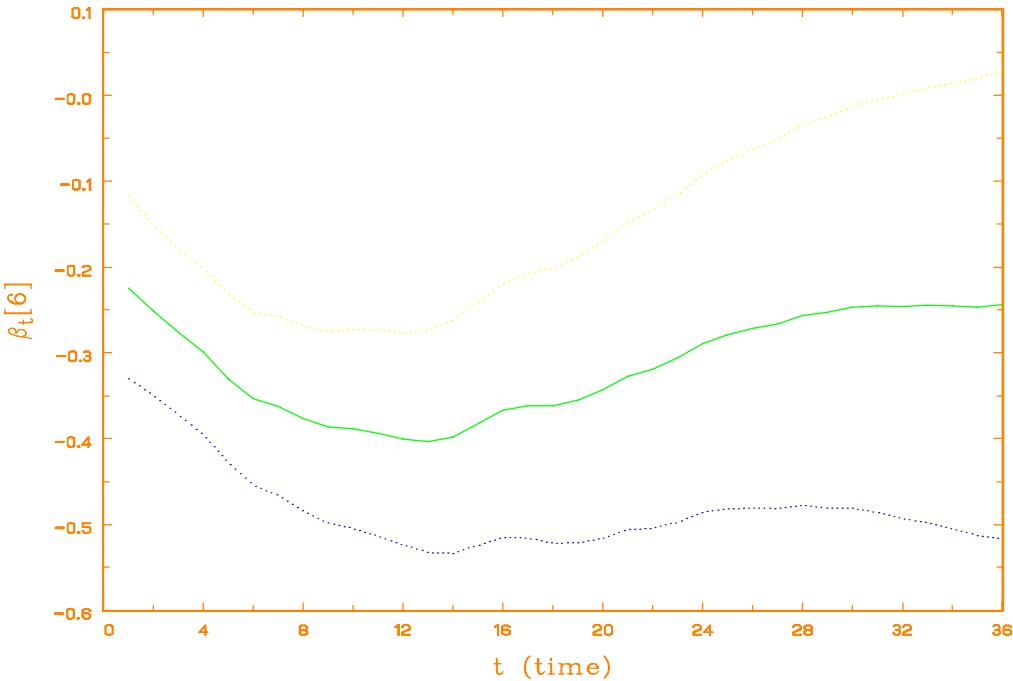


Abbildung 6.42 Geschätzter Altersseffekt zur Kovariablen A2 mit Konfidenzband.

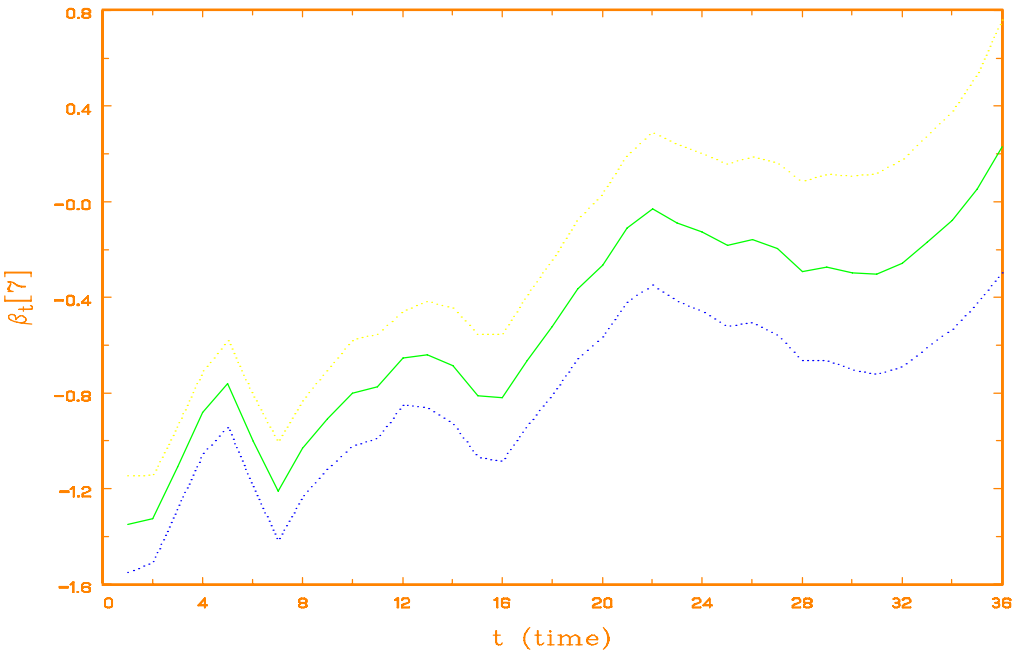


Abbildung 6.43 Geschätzter Altersseffekt zur Kovariablen A3 mit Konfidenzband.

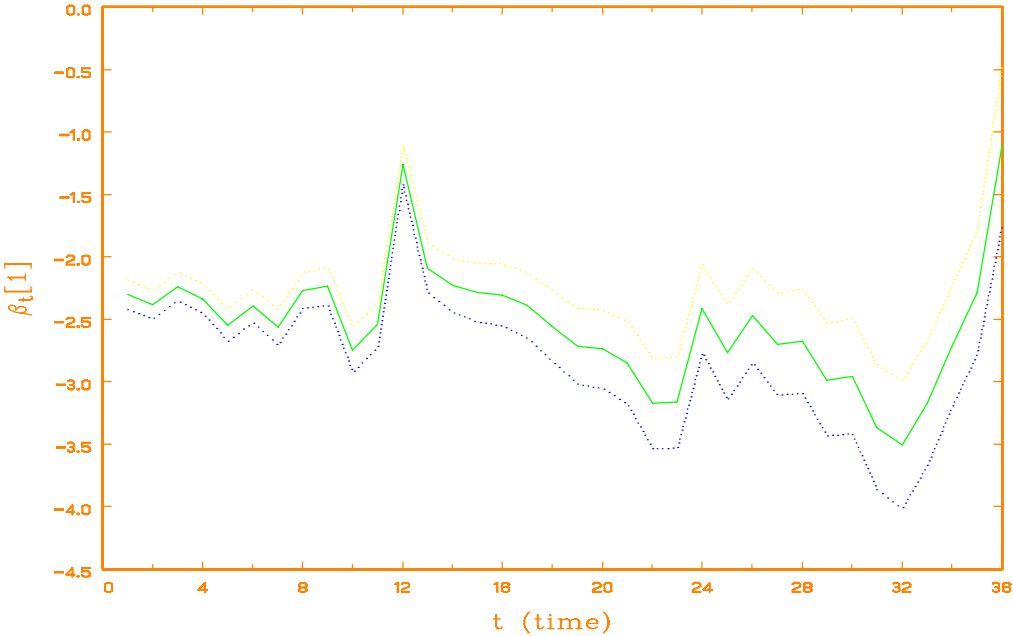


Abbildung 6.44 Geschätzter Baseline-Parameter $\tau_t = \beta_{1,t}$ mit Konfidenzband.

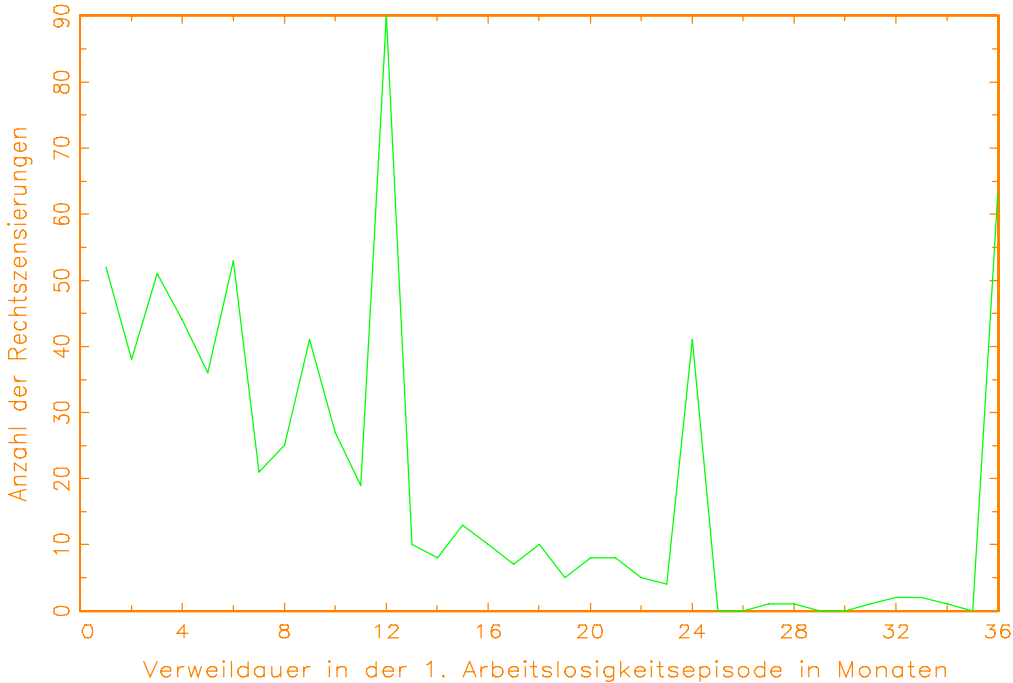


Abbildung 6.45 Anzahl der zensierten Abgänge aus der ersten Arbeitslosigkeitsepisode.

Die Abbildung 6.46 zeigt exemplarisch zwei Hazardraten von Personen, die, bis auf die Kovariable Geschlecht, jeweils der Referenzkategorie der Kovariablen angehören. Sei U_M ein Mann und U_F eine Frau. Dann ist in diesem Fall $\mathbf{x}_{U_M} = (0,0,1,0,0,0)'$ und $\mathbf{x}_{U_F} = \mathbf{0}_6$. Insbesondere sind die Werte der Hazardraten für diese Frauen die über (6.19) nichtlinear transformierten Werte der geschätzten Baseline-Parameter $\tau_{t|36}$.

Bis zum 9. Monat ist die Hazardrate für Männer signifikant über der von Frauen. Anschließend kehrt sich die Situation um. Die Hazardrate für Frauen liegt über der der Männer, allerdings nicht signifikant. Das hohe Ansteigen gegen Ende ist künstlich durch den Datensatz hervorgerufen: Alle Arbeitslosigkeitsdauern von mehr als 36 Monaten wurden abgeschnitten. Das heißt, die Arbeitslosigkeitsdauer wurde in einem solchen Fall auf 36 gesetzt und als zensiert markiert.

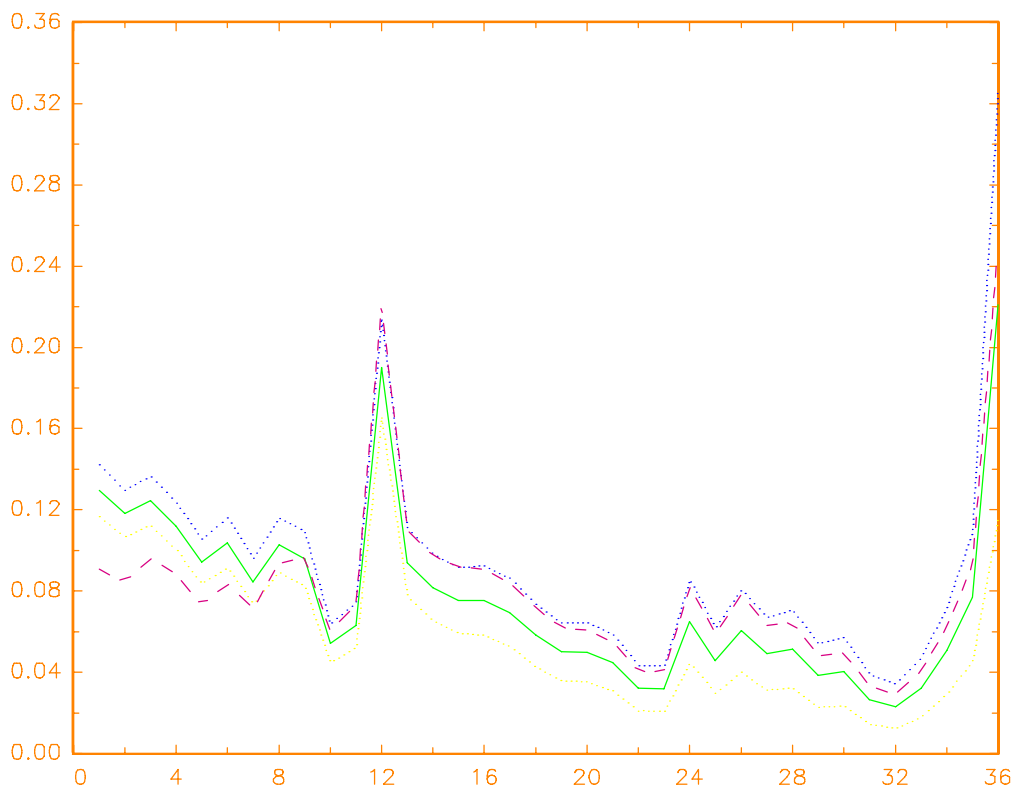


Abbildung 6.46: Geschätzte Hazardraten für ——— Männer und - - - - Frauen, die sich bei den Kovariablen Alter und Nationalität in der Referenzgruppe befinden.

6.5.2 Mehr-Zustands-Analyse

Bei der folgenden Untersuchung berücksichtigen wir nicht nur die Verweildauer in der Arbeitslosigkeit, sondern auch den Zustand, in den die Person anschließend wechselt. Dabei unterscheiden wir drei Abgangszustände aus der Arbeitslosigkeit:

- 1.) Abgang in eine Vollzeittätigkeit ($j = 1$).
- 2.) Abgang in eine Teilzeittätigkeit ($j = 2$).
- 3.) Abgang in einen sonstigen Zustand ($j = 3$).

In dem Abgangszustand $j = 3$ sind die Erwerbskategorien 3 und 5 bis 9 aus Teilabschnitt 1.1.5 zusammengefaßt. Zur Analyse verwenden wir im Beobachtungsmodell das dynamische multinomiale Logit-Modell (3.14) mit $\bar{J} = 3$ Kategorien in obiger Kodierung:

$$\lambda_j(t|\mathbf{x}_U) = \frac{\exp(\eta_{tUj})}{1 + \sum_{i=1}^3 \exp(\eta_{tUi})}, \quad j = 1, 2, 3, \quad t = 1, \dots, \bar{T} = 36, \quad U = 1, \dots, \bar{U} = 2911. \quad (6.20)$$

Der abgangsspezifische lineare Prädiktor $\eta_{tUj} = \tau_j + \mathbf{x}'_U \beta_j$ enthält den abgangs- und zeitabhängigen Baseline-Parameter τ_j und die abgangs- und zeitvariierenden Kovariableneffekte $\beta_j = (\beta_{2,j}, \beta_{3,j}, \beta_{4,j}, \beta_{5,j}, \beta_{6,j}, \beta_{7,j})'$ zum Kovariablenvektor \mathbf{x}_U wie in (6.19). Im Übergangsmodell verwenden wir wieder den RW1.

Über den Programmablauf bei der Posteriori-Modus-Schätzung informiert die Tabelle 6.15. In der Abbildung 6.47 sind die geschätzten Baseline-Parameter aufgetragen. Für die Abgangsarten Vollzeit- und Teilzeitbeschäftigung nimmt der Baseline-Parameter einen glatten Verlauf mit leichten Schwankungen. Die Spitzen nach 12, 24 und 36 Monaten, die wir in der Abbildung 6.44 im Ein-Zustands-Fall gesehen haben, sind deutlich beim geschätzten Baseline-Parameter für die Abgangsart Sonstiges zu erkennen. Weil in der Kategorie Sonstiges auch die zensierten Abgänge (0: keine Angabe beziehungsweise missing value) enthalten sind, bestätigt die Mehr-Zustands-Analyse unsere Erkenntnis über die Herkunft der Peaks aus Teilabschnitt 6.5.1.

Arbeitslosendaten, Mehr-Zustands-Fall: Dynamisches multinomiales Logit-Modell mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.015720532	0.015613662	704	$1 \cdot 10^{-10}$	88	$2 \cdot 10^{-4}$
0.003308079	0.002190734				
0.47666125	0.48087406				
0.009574057	0.009421104				
0.021018945	0.021443762				
0.002666955	0.001351732				
0.000886462	0.000537961				
0.000988374	0.000840554				
0.015632734	0.076161411				
0.005439112	0.005378416				
0.014106136	0.013465303				
0.016762889	0.020082894				
0.002028842	0.001564515				
0.004346374	0.003360897				
0.069875511	0.073082735				
0.006755622	0.006656125				
0.006918950	0.004419145				
0.002042488	0.001185151				
0.002304175	0.001360718				
0.007345425	0.005322596				
0.028181438	0.022290397				

Tabelle 6.15: Hyperparameterschätzung und Programmablauf von PME.ST.

Die geschätzten Effekte, Ost-Deutscher zu sein, sind in der Abbildung 6.48 angegeben. Für die Abgangsarten Voll- und Teilzeit sind die Effekte nahezu zeitkonstant und signifikant positiv. Sie verlaufen innerhalb eines Konfidenzbandes. Der Effekt für die Abgangsart Sonstiges ist zu Beginn positiv. Ab dem 8. Monat zeigt er fallenden Tendenz und ist nach einem Jahr signifikant negativ. Als Ergebnis bleibt festzuhalten, daß Ost-Deutsche auf dem Arbeitsmarkt signifikant bessere Chancen haben als Ausländer (West), wenn es darum geht, eine Vollzeit- oder Teilzeittätigkeit zu bekommen. Für den Verlauf des Effekts in die

Abgangsart Sonstiges mögen einige Sondereinflüsse der Deutschen Wiedervereinigung eine Rolle gespielt haben. Damit haben wir eine weitaus detailliertere Analyse des Nationalitätseffektes N2 vorliegen als mit der Abbildung 6.39 im Ein-Zustands-Fall.

In der Abbildung 6.49 sind die Geschlechtseffekte dargestellt. Um eine Vollzeittätigkeit zu bekommen, haben Männer signifikant bessere Chancen als Frauen. Umgekehrt ist die Situation bei der Abgangsart Teilzeittätigkeit. Dies mag auch darauf zurückzuführen sein, daß prozentual mehr Frauen eine Teilzeitbeschäftigung suchen als Männer. Der Effekt für die Abgangsart Sonstiges verläuft mit leichten Schwankungen stets negativ.

Die Abbildung 6.50 zeigt die Alterseffekte zur Altersgruppe A3. Die Effekte für die Abgangsarten Voll- und Teilzeit verlaufen fast parallel, nahezu zeitkonstant und signifikant negativ, wie man es auch erwarten würde. Der Effekt für den Abgang in den Zustand Sonstiges liegt auf einem höheren Niveau. Er ist zu Beginn signifikant negativ, steigt im Verlauf der Arbeitslosigkeit an und ist bei einer Arbeitslosigkeitsdauer von 3 Jahren sogar signifikant positiv. Eine Erklärung für diesen Sachverhalt ist, daß die über 50-jährigen Arbeitslosen im Vergleich zur Referenzgruppe zwar geringere Chancen haben, eine Vollzeit- oder Teilzeittätigkeit zu erhalten, jedoch bei längerer Arbeitslosigkeitsdauer in die Rente, die Pension oder den Vorruhestand wechseln.

In den Abbildungen 6.51 und 6.52 sind exemplarisch die Hazardraten von Frauen und von Männern geplottet, die zur Gruppe der Ausländer (West) und der zwischen 31- und 40-jährigen gehören. Die Abbildung 6.51 zeigt die Hazardraten für den Abgang in die Vollzeit. Dabei haben Männer zu Beginn eine signifikant höhere Hazardrate. Im weiteren Verlauf der Arbeitslosigkeit verringert sich die Differenz der Hazardraten. Die der Männer liegt jedoch immer noch über der der Frauen. Bei den Hazardraten für den Abgang in die Teilzeit ist die Situation genau umgekehrt: Die Hazardrate für die Frauen liegt signifikant über der der Männer. Die geschätzten Hazardraten korrespondieren mit den geschätzten Geschlechtseffekten in der Abbildung 6.49.

Im Ökonomischen Bereich werden in den letzten Jahren verstärkt Mehr-Zustands-Modelle zur Analyse der Arbeitslosigkeitsdauern eingesetzt (vergleicher etwa Narendranathan und Stewart, 1993, und Pudney und Thomas, 1995). Narendranathan und Stewart (1991) geben einen Test an, ob die „proportional hazards“-Eigenschaft der zustandsspezifischen Hazardrate erfüllt ist.

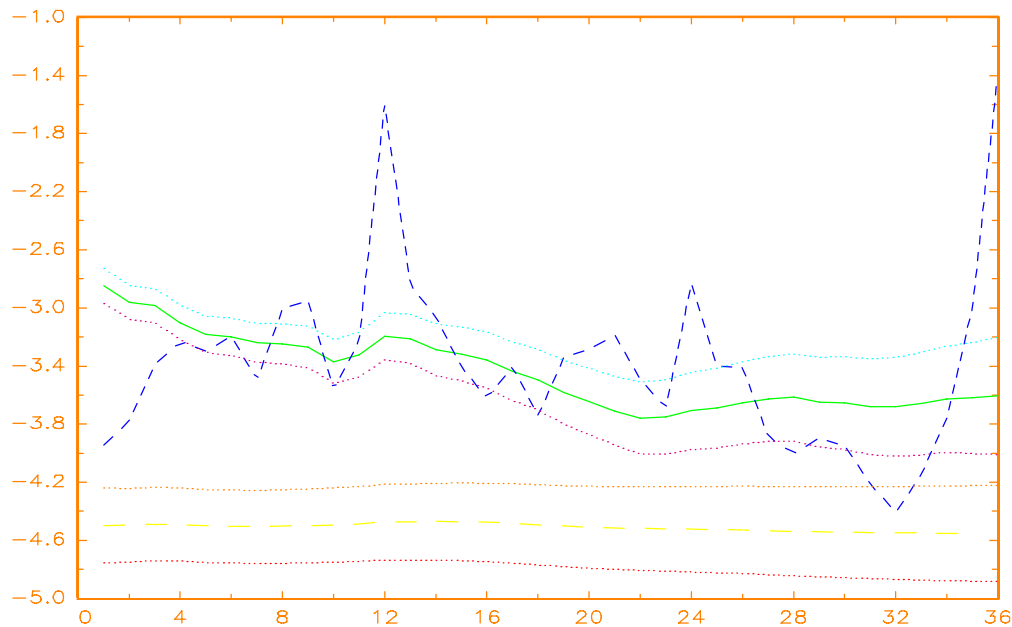


Abbildung 6.47: Geschätzte Baseline-Parameter für die Abgangsart Vollzeit —, die Abgangsart Teilzeit - - - und die Abgangsart Sonstiges Mehr-Zustands-Modell, 1. Arbeitslosigkeitsepisode.

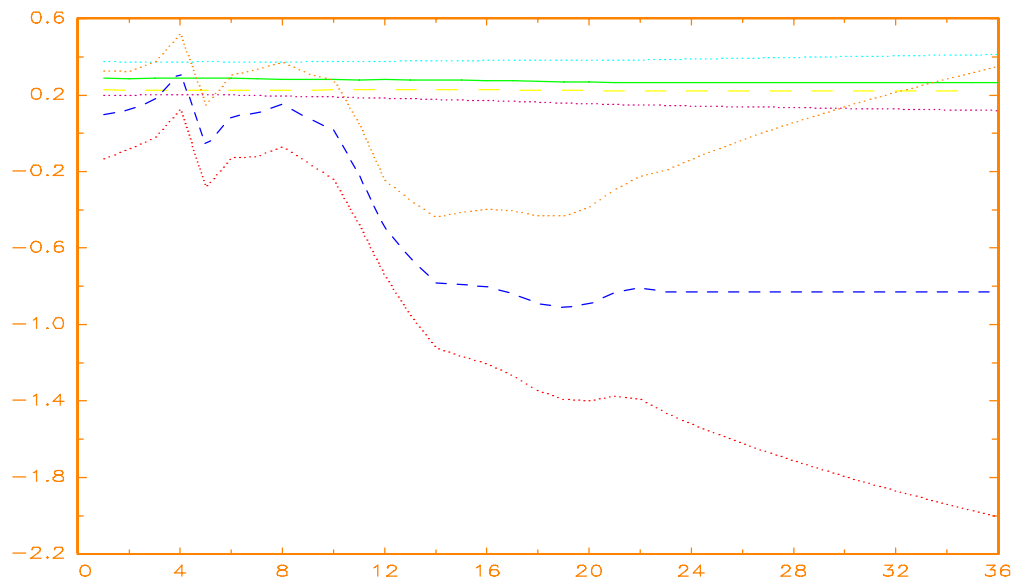


Abbildung 6.48: Geschätzter Nationalitätseffekt N2 (Ost-Deutsche) für die Abgangsart Vollzeit —, die Abgangsart Teilzeit - - - und die Abgangsart Sonstiges Mehr-Zustands-Modell, 1. Arbeitslosigkeitsepisode.

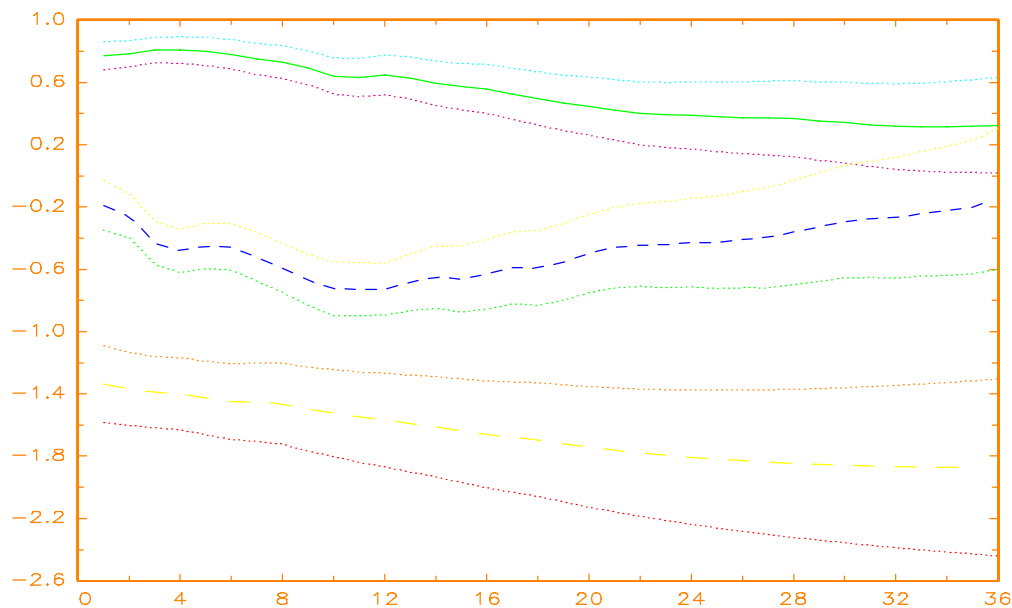


Abbildung 6.49: Geschätzter Geschlechtseffekt für die Abgangsart Vollzeit ———, die Abgangsart Teilzeit - - - - und die Abgangsart Sonstiges - - - - - . Mehr-Zustands-Modell, 1. Arbeitslosigkeitsepisode.

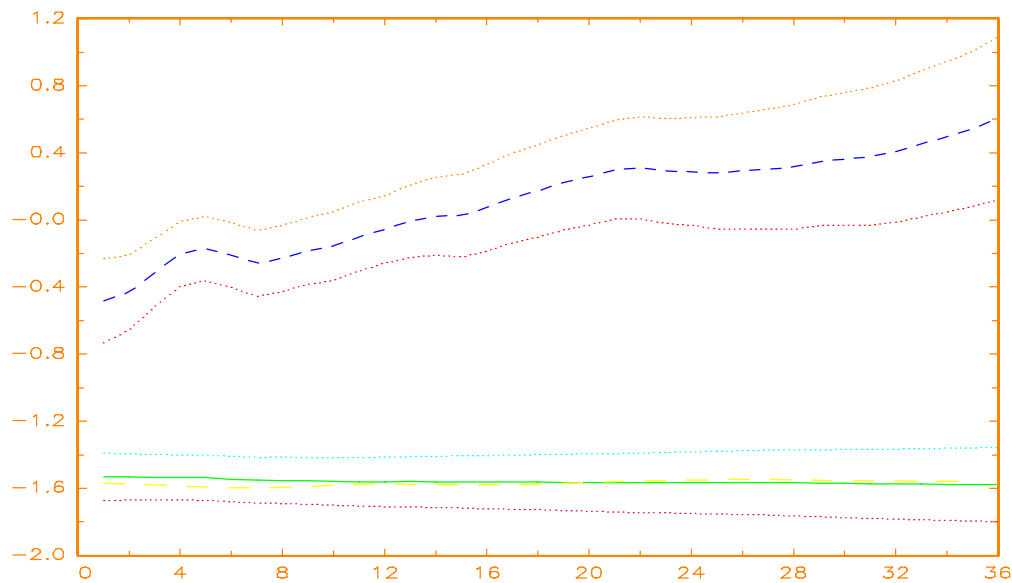


Abbildung 6.50: Geschätzter Alterseffekt A3 für die Abgangsart Vollzeit ———, die Abgangsart Teilzeit - - - - und die Abgangsart Sonstiges - - - - - . Mehr-Zustands-Modell, 1. Arbeitslosigkeitsepisode.

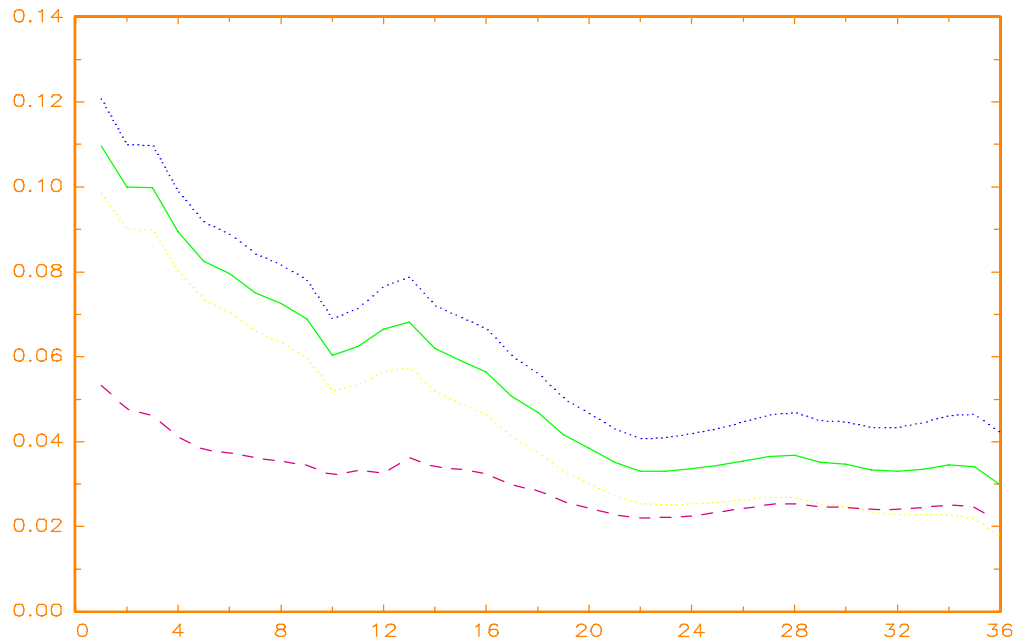


Abbildung 6.51: Geschätzte Hazardraten für die Abgangsart Vollzeit: — Männer, Frauen — — — . Die Werte der übrigen Kovariablen liegen in den Referenzkategorien. Mehr-Zustands-Modell, 1. Arbeitslosigkeitsepisode.

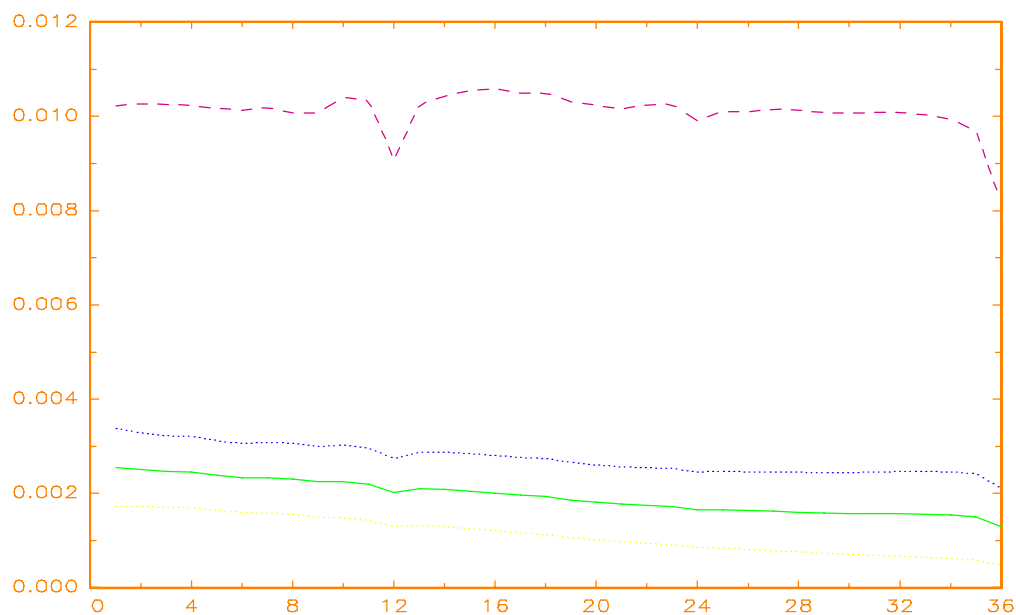


Abbildung 6.52: Geschätzte Hazardraten für die Abgangsart Teilzeit: — Männer, Frauen — — — . Die Werte der übrigen Kovariablen liegen in den Referenzkategorien. Mehr-Zustands-Modell, 1. Arbeitslosigkeitsepisode.

6.5.3 Mehr-Episoden-Analyse

In den vorhergehenden beiden Teilabschnitten 6.5.1 und 6.5.2 haben wir die erste Arbeitslosigkeitsepisode untersucht. Im vorliegenden Datensatz gab es 816 Personen, die ein zweites Mal arbeitslos wurden. Diese zweite Arbeitslosigkeitsepisode wollen wir nun analysieren unter der Voraussetzung, daß die Baseline-Parameter und die Kovariableneffekte episodenspezifisch sind. Damit können wir zur Bestimmung des Posteriori-Modus-Schätzers die penalisierte Loglikelihood getrennt für jede Episode maximieren. Für die Untersuchung der zweiten Arbeitslosigkeitsepisode ist also nur $\text{PL}_2(\alpha_{\bar{T}^*}^2)$ aus (4.66) zu maximieren. Als Beobachtungsmodell verwenden wir, womit auch die Beobachtungsloglikelihood $p(y_{tU}^2 | y_{(t-1)^*}^{2*}, y_t^1, x_{t^*}^{2*}, r_{t^*}^{2*}, \alpha_t^2)$ in (4.66) festgelegt ist, das dynamische multinomiale Logit-Modell:

$$\lambda_j^2(t|x_U, x_U^2) = \frac{\exp(\eta_{tUj}^2)}{1 + \sum_{i=1}^3 \exp(\eta_{tUi}^2)}, \quad j = 1, 2, 3, \quad t = 1, \dots, \bar{T} = 36, \quad U = 1, \dots, \bar{U} = 816. \quad (6.21)$$

t gibt dabei die Verweildauer- und nicht die Kalenderzeit an. Das bedeutet, daß sämtliche Aussagen, die die Zeit betreffen, für jede Person auf den Monat des Beginns ihrer zweiten Arbeitslosigkeit bezogen sind. $j \in \{1, 2, 3\}$ bezeichnet wie im Mehr-Zustands-Fall in Teilabschnitt 6.5.2 die Abgangsart. Der abgangsspezifische lineare Prädiktor der Episode $k = 2$, $\eta_{tUj}^k = \tau_{tj}^k + (x_U^{k*})' \beta_{tj}^k$, enthält den abgangs- und zeitabhängigen Baseline-Parameter τ_{tj}^k und die abgangs- und zeitvariierenden Kovariableneffekte $\beta_{tj}^k = (\beta_{2,tj}^k, \beta_{3,tj}^k, \beta_{4,tj}^k, \beta_{5,tj}^k, \beta_{6,tj}^k, \beta_{7,tj}^k, \beta_{8,tj}^k, \beta_{9,tj}^k)'$ der $k = 2$ -ten Episode zum Kovariablenvektor $x_U^{2*} = (x_U', (x_U^2)')'$ mit x_U wie in (6.19). Im Kovariablenvektor x_U^2 sind die metrischen, exogenen Einflußgrößen V = Verweildauer in der ersten Arbeitslosigkeitsepisode, gemessen im Monaten, und E = Dauer der Erwerbstätigkeit zwischen der ersten und der zweiten Arbeitslosigkeitsepisode in Monaten, zusammengefaßt. Die zugehörigen Effekte werden mit $\beta_{8,tj}^k$ und $\beta_{9,tj}^k$ bezeichnet. Im Übergangsmodell verwenden wir wieder für jeden zu schätzenden Parameter (Baseline-Parameter und Kovariableneffekte) den RW1.

Über den Programmablauf bei der Posteriori-Modus-Schätzung informiert die Tabelle 6.16.

Arbeitslosendaten, Mehr-Episoden-Fall: Dynamisches multinomiales Logit-Modell mit RW1					
diag($Q^{(0)}$)	diag(\hat{Q})	FS-Iter.	FS_acc.	EM-Iter.	EM_acc.
0.009860985	0.001654681	2264	$1 \cdot 10^{-10}$	283	$2 \cdot 10^{-4}$
0.003458349	0.002030443				
0.48530109	0.28271301				
0.011119422	0.006581492				
0.018600105	0.003948636				
0.002172643	0.000969939				
0.001005576	0.002084314				
0.001000391	0.001958844				
0.001003179	0.001013705				
0.012227553	0.009243484				
0.018678683	0.075937475				
0.006700593	0.076803733				
0.003461965	0.001268235				
0.014727257	0.074636674				
0.004798339	0.002110119				
0.007978346	0.006120479				
0.020043711	0.012891041				
0.009151900	0.16563200				
0.008002797	0.002024460				
0.012674152	0.008835216				
0.013202068	0.045814281				
$1 \cdot 10^{-6}$	$2.7901 \cdot 10^{-6}$				
$1 \cdot 10^{-6}$	$9.8529 \cdot 10^{-7}$				
$1 \cdot 10^{-6}$	$3.6010 \cdot 10^{-6}$				
$1 \cdot 10^{-6}$	$4.1612 \cdot 10^{-7}$				
$1 \cdot 10^{-6}$	$1.3307 \cdot 10^{-6}$				
$1 \cdot 10^{-6}$	$4.3238 \cdot 10^{-5}$				

Tabelle 6.16: Hyperparameterschätzung und Programmablauf von PME.ST.

Die Abbildung 6.53 enthält die für die zweite Arbeitslosigkeit im Verweildauerfall geschätzten Baseline-Parameter. Wie im Mehr-Zustands-Fall nehmen die Schätzungen für die Abgangsarten Voll- und Teilzeit einen glatten Verlauf mit leichten Schwankungen. Auch bei der zweiten Arbeitslosigkeitsepisode treten für den Abgang in die Kategorie Sonstiges die charakteristischen Spitzen bei 12 und 24 Monaten auf. Der Peak bei 36 Monaten fehlt, weil nur 11 von den 816 Verweildauern in der zweiten Arbeitslosigkeit 36 Monate oder länger sind.

Für die Nationalitäts-, Geschlechts- und Alterseffekte ergeben sich im Vergleich zur ersten Episode in unserer Mehr-Zustands-Analyse von Teilabschnitt 6.5.2 keine Änderungen. Deshalb sei an dieser Stelle auf die Abbildungen 6.48, 6.49 und 6.50 verwiesen. Insbesondere haben auch bei der zweiten Arbeitslosigkeitsepisode die Männer signifikant bessere Chancen als Frauen, eine Vollzeitbeschäftigung zu bekommen. Umgekehrt verhält es sich mit der Abgangsart Teilzeit. Bei der Interpretation ist jedoch zu beachten, daß möglicherweise bei weitem mehr Frauen als Männer nach einer Teilzeitarbeit suchen. Ebenso verlaufen die Schätzungen für die Alterseffekte A_3 in die Abgänge Voll- und Teilzeit signifikant negativ, während sich für die Abgangsart Sonstiges ein vom Negativen ins Positive wachsender Trend ergibt. Offenbar warten auch bei der zweiten Arbeitslosigkeit die über 50-jährigen darauf, in die Rente, Pension oder den Vorruhestand gehen zu können. In der Abbildung 6.54 sind die Effekte der metrischen Kovariablen „Dauer der ersten Arbeitslosigkeit“ aufgetragen. Für die Abgangsarten Voll- und Teilzeit sind sie zeitkonstant (die entsprechenden Komponenten von \hat{Q} wurden sehr klein geschätzt) und signifikant negativ. Das bedeutet: Je länger die erste Arbeitslosigkeit gedauert hat, desto geringer sind die Chancen, im Anschluß an die zweite Arbeitslosigkeit eine Voll- oder Teilzeitbeschäftigung zu bekommen, ganz gleich, wie lange die zweite Arbeitslosigkeitsepisode schon andauert. Die Schätzung für den Abgang in die Kategorie Sonstiges liegt beim Nullniveau. Für diese Abgangsart ist es daher unerheblich, ob eine erste Arbeitslosigkeitsepisode vorlag und wie lange sie gedauert hat.

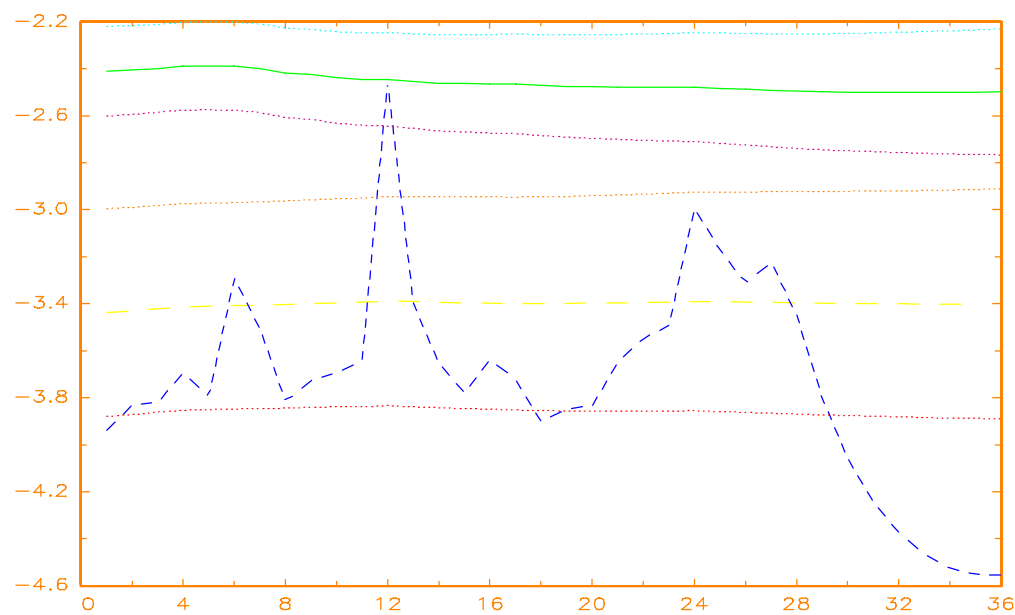


Abbildung 6.53: Geschätzter Baseline-Parameter für die Abgangsart Vollzeit ———, die Abgangsart Teilzeit - - - - und die Abgangsart Sonstiges - - - - - . Mehr-Episoden-Modell, 2. Arbeitslosigkeitsepisode. Abszisse: Monate.

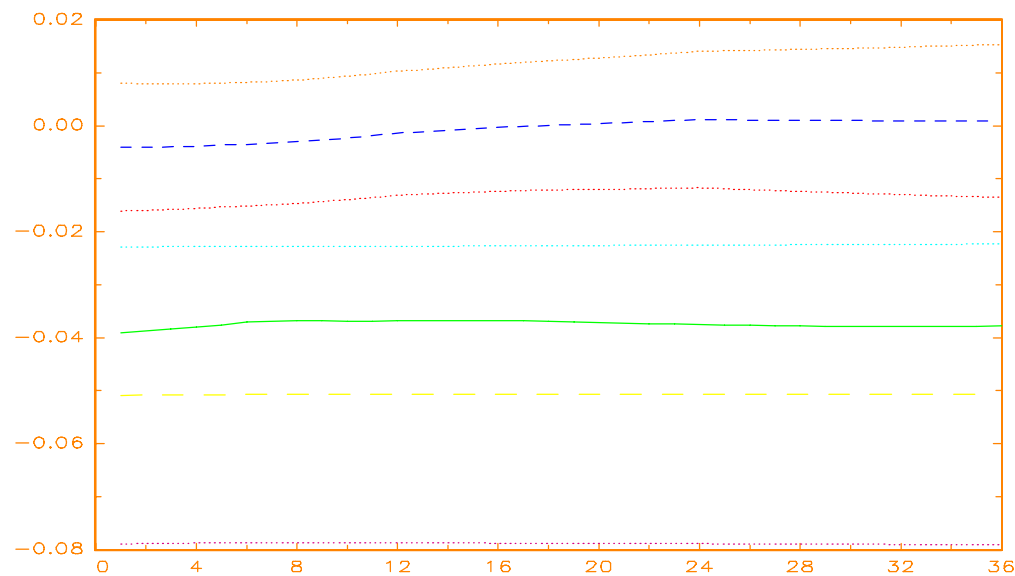


Abbildung 6.54: Geschätzte Effekte zur Dauer der ersten Arbeitslosigkeitsepisode für die Abgangsart Vollzeit ———, die Abgangsart Teilzeit - - - - und die Abgangsart Sonstiges - - - - - . Mehr-Episoden-Modell, 2. Arbeitslosigkeitsepisode. Abszisse: Zeit in Monaten.

Die Abbildung 6.55 zeigt die geschätzten Effekte der metrischen Kovariable „Dauer der Erwerbstätigkeit zwischen der ersten und zweiten Arbeitslosigkeitsepisode“. Die Effekte für die Abgangsarten Voll- und Teilzeit sind zeitkonstant und negativ. Möglicherweise spielt dabei die Tatsache eine Rolle, daß die betreffenden Personen bereits einmal arbeitslos waren. Die Schätzung für den Abgang in den Bereich Sonstiges ist zeitvariierend. Sie ist zu Beginn der zweiten Arbeitslosigkeit negativ und wächst im weiteren Verlauf an. Nach 7 Monaten wird sie positiv und bleibt weiterhin im positiven Bereich. Mit zunehmender zweiter Arbeitslosigkeitsdauer (ab dem 7. Monat) hat die Dauer der Erwerbstätigkeit zwischen der ersten und zweiten Arbeitslosigkeitsepisode einen wachsenden positiven Einfluß und damit stärkeres positives Gewicht auf den Abgang aus der zweiten Arbeitslosigkeit in den Zustand „Sonstiges“. Ab dem 22. Monat nimmt der Einfluß wieder ab, bleibt aber positiv.

Verwendet man für die Zeit t in (6.21) nicht die Verweildauer- sondern die Kalenderzeit, so bekommt man eine kalenderzeitorientierte Schätzung der Baseline-Parameter und der Effekte für die zweite Arbeitslosigkeitsepisode. Das 1. Monat bezeichnet dabei den Januar 1983. Nach dem 108. Monat, das ist der Dezember 1991, ist keine Person im Datensatz aus der zweiten Arbeitslosigkeitsepisode abgegangen. Deshalb sind in der Abbildung 6.56, die die geschätzten Baseline-Parameter bei Kalenderzeit t darstellt, nur die ersten 108 Monate ab Januar 1983 aufgetragen. In der Schätzung des Baseline-Parameters für die Abgangsart Sonstiges finden sich wieder die jährlichen Spitzen nach 12, 24, 36, usw. Monaten. Ferner ist für diese Abgangsart ein ansteigender Trend zu verzeichnen. Damit liegt nach dieser Schätzung zum Ende der 80er Jahre hin eine steigende Tendenz vor, nach der zweiten Arbeitslosigkeit in die Kategorie Sonstiges zu wechseln. Eine mögliche Erklärung hierfür ist die von der Bundesregierung in der Mitte der 80er Jahre beschlossene Vorruhestandsregelung. Für den 1. Monat haben die Schätzungen den Wert Null, weil im Januar 1983, das ist der Beginn des Beobachtungszeitraumes, noch niemand seine zweite Arbeitslosigkeitsepisode verlassen hat. Die Schätzungen für die Abgangsarten Voll- und Teilzeit bleiben nach einem Jahr auf etwa dem selben Niveau.

Für eine Interpretation der kalenderzeitorientierten Effekte ist das verwendete Modell zu wenig komplex. Dies muß einer größer angelegten ökonomischen Untersuchung vorbehalten bleiben.

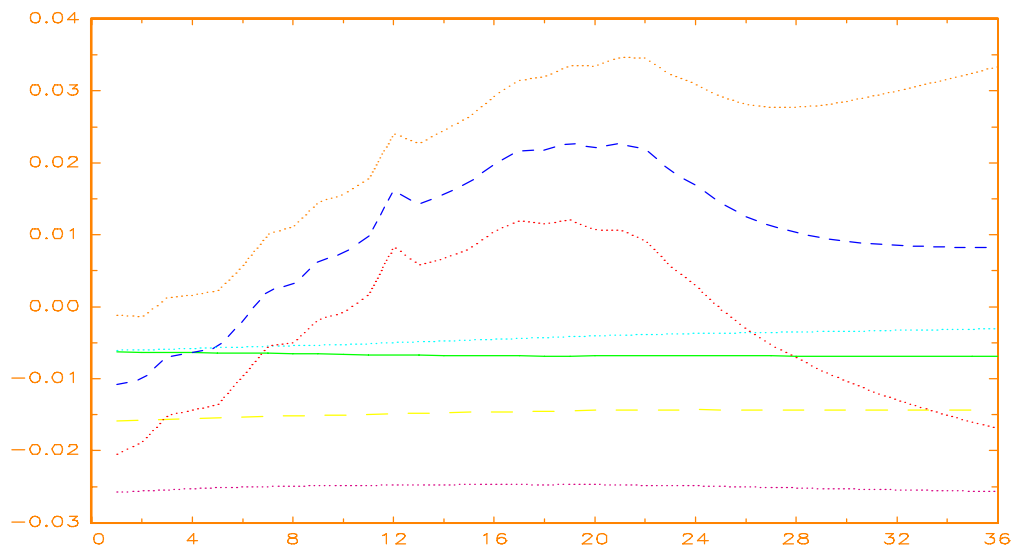


Abbildung 6.55: Geschätzte Effekte zur Dauer der Erwerbstätigkeit zwischen 1. und 2. Arbeitslosigkeit für die Abgangsart Vollzeit ———, die Abgangsart Teilzeit — — — — und die Abgangsart Sonstiges - - - - - . Mehr-Episoden-Modell, 2. Arbeitslosigkeitsepisode. Abszisse: Zeit in Monaten.

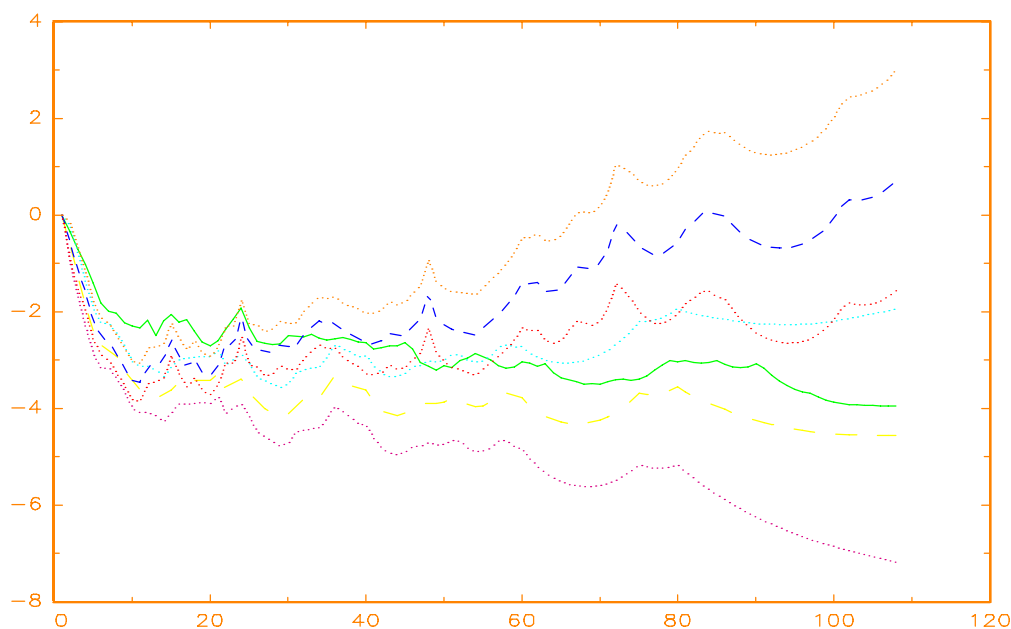


Abbildung 6.56: Geschätzte Baseline-Parameter für die Abgangsart Vollzeit ———, die Abgangsart Teilzeit — — — — und die Abgangsart Sonstiges - - - - - . Mehr-Episoden-Modell, 2. Arbeitslosigkeitsepisode. Abszisse: Kalenderzeit t in Monaten.

Die verwendeten Modelle dienen hauptsächlich zur Darstellung der methodischen Möglichkeiten. Sie können einen ersten Anhaltspunkt geben über die Wirkung und den Einfluß der einbezogenen Kovariablen. Für ökonomisch fundierte Aussagen sind umfangreichere und differenziertere Modelle notwendig, wie sie im Competing-Risks-Kontext beispielsweise in Narendranathan und Stewart (1993) und Pudney und Thomas (1995) besprochen werden.

Gleichzeitig ergibt sich ein Trade-off zwischen numerischem Aufwand und Umfang des Modells. Im Ein-Episoden-Ein-Zustands-Fall in Teilabschnitt 6.5.1 war der numerische Aufwand vergleichsweise gering. Die Rechenzeiten waren das für Mehr-Zustands-Modell höher, die Schätzergebnisse jedoch detaillierter und aufschlußreicher. Manche Zusammenhänge ließen sich überhaupt erst anhand der Mehr-Zustands-Analyse erkennen.

Dritte und höhere Arbeitslosigkeitsepisoden können mit dem hier beschriebenen methodischen und algorithmischen Instrumentarium in analoger Weise untersucht werden. Aus Mangel an Fallzahlen im vorhandenen Datensatz wurde dies nicht durchgeführt.

7 Zusammenfassung und Ausblick

Im Rahmen dieser Arbeit wurden zeitdiskrete Modelle zur Ereignisanalyse mit dynamischen Effekten vorgestellt. Dabei sind auch Modelle mit nicht proportionalen Hazards zugelassen. Wir haben sie in Zustandsraumform angegeben. In diesem allgemeinen Modellrahmen wurde das Posteriori-Modus-Schätzkonzept zur Bestimmung zeitabhängiger Effekte eingesetzt. Zur numerisch effizienten Schätzung haben wir den bekannten (KFG)-Algorithmus zum (LGKFG) modifiziert und diesen wiederum iteriert (IGKFG). Außerdem haben wir ein Schätzverfahren (D-IGKFG) angegeben, das den Fall einer diffusen beziehungsweise nichtinformativen Start-Priori-Verteilung numerisch stabil und effizient behandelt.

Verallgemeinerungen und Weiterentwicklungen der hier angegebenen Konzepte sind in verschiedene Richtungen möglich.

- 1.) Zur Behandlung von unbeobachteter Populationsheterogenität kann es sinnvoll sein, zusätzlich zufällige Effekte in die Modellierung miteinzubeziehen. Für den Bereich des Gibbs-Sampling ist dies in Knorr-Held (1995) bereits geschehen.
- 2.) Ein systematischer Vergleich von Posteriori-Mean-Schätzung (Gibbs Sampling) und Posteriori-Modus-Schätzung ist wichtig, um die erzielten Resultate weiter bewerten und einordnen zu können.
- 3.) Der Bereich Modelldiagnose und Modellcheck kann noch weiter ausgebaut werden. Gamerman und West (1987) schlagen die Verwendung von Bayes-Faktoren vor. Kass und Raftery (1994, Teilabschnitt 4.1.1) besprechen unter anderem die Laplace-Methode zur Bestimmung von Bayes-Faktoren. Mit ihrer Formel (4) geben sie eine interessante Möglichkeit, Bayes-Faktoren unter Verwendung des Posteriori-Modus-Schätzers zu berechnen.
- 4.) Die Einbeziehung von zeitabhängigen Kovariablen in die hier behandelten dynamischen Modell ist vom theoretischen Standpunkt aus möglich. Damit kann man dynamische Effekte in Verbindung mit zeitvariierenden Kovariablen schätzen. Weitere praktische Untersuchungen sind nötig, um für diesen Fall das Laufzeitverhalten der hier

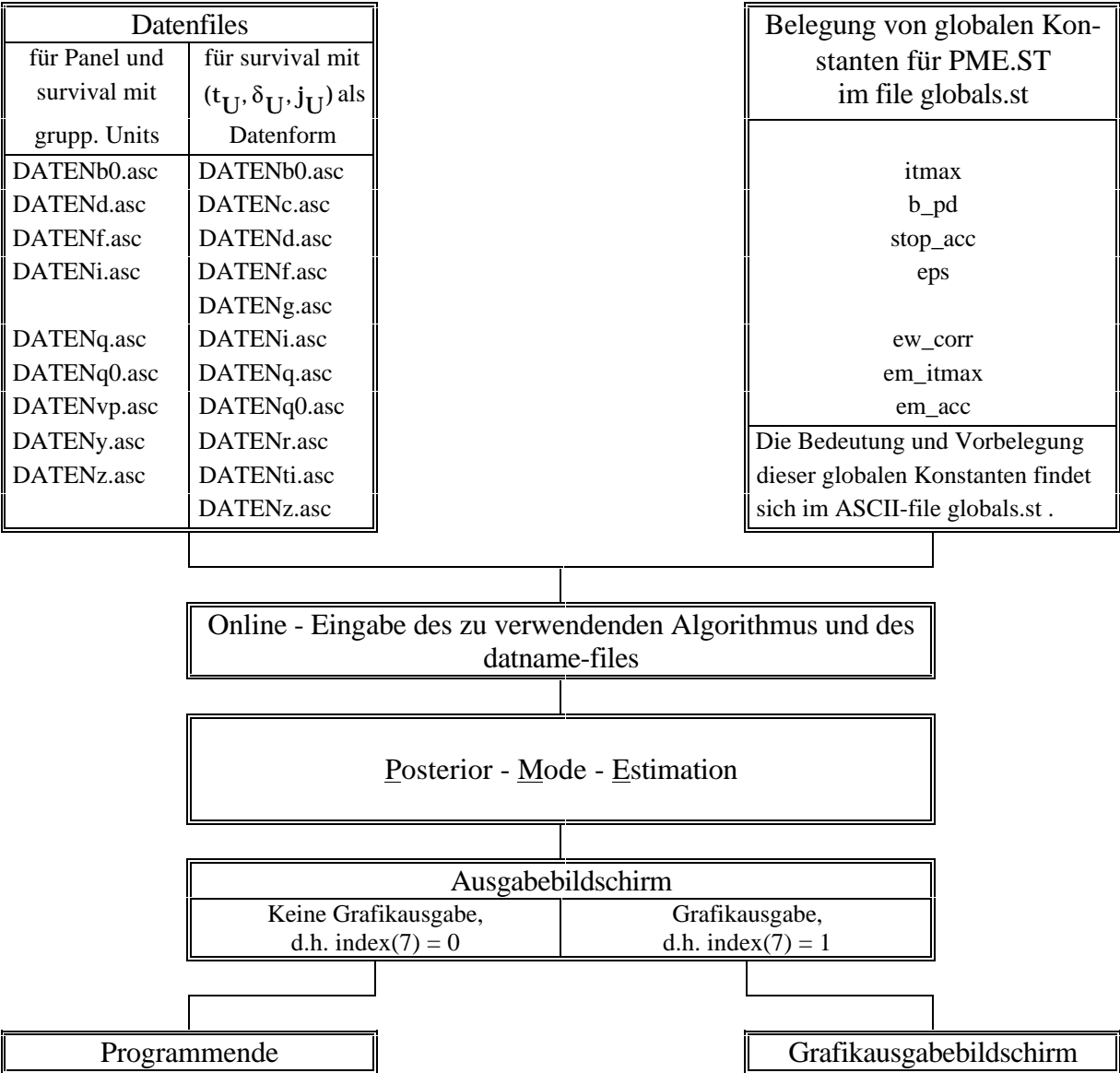
besprochenen Schätzprozeduren zu eruieren. Außerdem muß auf eventuell auftretende Identifikationsprobleme geachtet werden.

- 5.) Die in Fahrmeir und Wagenpfeil (1995a) enthaltene Simulationsstudie gibt einen Hinweis darauf, daß die in Kapitel 6 angegebenen punktweisen Konfidenzbänder „eher zu breit“ sind, das heißt man befindet sich auf der sicheren Seite. Interessant wären weitere theoretische Untersuchungen zu den asymptotischen Eigenschaften des Posteriori-Modus-Schätzers.

Anhang

A Programmbeschreibung zu PME.ST

Das folgende Schema veranschaulicht den strukturellen Programmablauf von PME.ST . Die einzelnen Programmpunkte werden im Anschluß näher erläutert.



Ablaufdiagramm zu PME.ST

Bemerkungen zum Ablaufdiagramm von PME.ST:

- Datenfiles:

Wie auch im Ablaufdiagramm zu PME.ST wird im folgenden angenommen, der Datensatzname sei DATEN. Dann erwartet PME.ST vom Benutzer, daß er die nachstehenden ASCII-files, im Text als Datenfiles bezeichnet, zur Verfügung stellt:

File	enthält
DATENb0.asc	(Start-) a_0 aus der Anfangsbedingung des Übergangsmodells, falls informative Start-Priori..
DATENC.asc	Enthält für jede Beobachtungseinheit (= Unit) den Zensierungsindikator. Dieses File wird nur für Survival-Daten in der Form (t_U, δ_U, j_U) benötigt.
DATEND.asc	Wird im PME.ST als datname-file bezeichnet und enthält den oder die Datensatznamen, hier im Beispiel also DATEN.
DATENf.asc	Übergangsmatrix aus dem Übergangsmodell.
DATENG.asc	Enthält den grid-Vektor, der die Zeitintervallängen in Basis-einheiten angibt.
DATENi.asc	Indexfile: (→ siehe Abschnitt: Belegung des Vektors <i>index</i> in PME.ST).
DATENq.asc	(Start-) Kovarianzmatrizen des Übergangsmodells.
DATENq0.asc	(Start-) Kovarianzmatrix aus der Anfangsbedingung des Übergangsmodells, falls informative Start-Priori.
DATENr.asc	Enthält für jede Beobachtungseinheit (= Unit) den Zustand, in den sie wechselt. Dieses File wird nur für Competing risks-Daten in der Form (t_U, δ_U, j_U) benötigt.
DATENti.asc	Enthält für jede Beobachtungseinheit (= Unit) den Zeitpunkt des Zustandswechsels. Dieses File wird nur für Survival-Daten in der Form (t_U, δ_U, j_U) benötigt.
DATENvp.asc	Für Normalverteilung: Enthält die Varianz-Kovarianzmatrizen des Beobachtungsmodells. Für Multinomialverteilung $M(n_U, \pi)$: Enthält n_U . Dieses File wird nur für Panel-Daten benötigt und im Survival-Fall mit gruppierten Beobachtungseinheiten.
DATENy.asc	Die beobachteten Werte. Dieses File wird nur für Panel-Daten benötigt und im Survival-Fall mit gruppierten Beobachtungseinheiten.
DATENZ.asc	Die Design-Matrix/-Matrizen.

Der Datensatzname muß ein maximal 6-stelliger und der Name des datname-files ein höchstens 8-stelliger alphanumerischer Ausdruck sein, wobei die erste Stelle mit einem Buchstaben zu belegen ist.

Survival-Daten können in zwei verschiedenen Formen für PME.ST aufbereitet werden:

1. Form: (für Panel-Daten und im Survival-Fall mit gruppierten Beobachtungseinheiten)

Im File DATENvp.asc steht n_{tU} mit n_{tU} = Mächtigkeit des risk set zum Zeitpunkt t für Gruppe/Beobachtungseinheit Nr. U (= Anzahl der Personen aus Gruppe U , die zum Zeitpunkt t noch unter Risiko stehen).

Im File DATENy.asc steht y_{tU} mit y_{tU} = (im Ein-Zustands-Fall) Anzahl der Personen aus Gruppe U , die zum Zeitpunkt t sterben, und y_{tUj} = (im Mehr-Zustands-Fall) j -te Komponente von $y_{tUj} \in \mathbb{R}^{\bar{J}}$, wobei \bar{J} = Anzahl der Zustände und y_{tUj} = Anzahl der Personen aus Gruppe U , die zum Zeitpunkt t in den Zustand j wechseln ($1 \leq j \leq \bar{J}$).

Diese Form ist besonders für die Gruppierung von Beobachtungseinheiten mit gleichen Kovariablen für alle Zeitpunkte geeignet, wenn $\bar{U} \geq \bar{T}$.

2. Form: (t_U, δ_U, j_U)

Bezeichnungen wie in Abschnitt 2.5.

Die j_U werden nur für den Competing risks-Fall benötigt.

Jeder Survival-Datensatz kann bis auf Permutation bezüglich der Beobachtungseinheiten eindeutig von einer in die andere Form gebracht werden.

- Belegung von globalen Konstanten für PME.ST im File globals.st:

Auszug aus globals.st:

```

itmax      = 50;                /* Maximum number of iterations for
                                Fisher-scoring */
b_pd       = 0.04;              /* Bound for positive definite - correction */
stop_acc   = 1e-10;             /* Stopping accuracy for stopping criterion */
eps        = 1.1108166e-16;     /* Machine - precision of Gauss V3.0 */
ew_corr    = 0;                 /* Eigenvalue - Correction:
                                0: no correction
                                1: simple eigenvalue - correction
                                2: sophisticated eigenvalue - correction */
em_itmax   = 400;               /* Maximum number of EM-type steps */

```

```

em_acc      = 0.001;          /* Stopping accuracy for stopping criterion
                                (EM-type algorithm)                */

```

- Algorithmen:

Bisher sind 4 Algorithmen zur Posteriori-Modus-Schätzung implementiert:

```

EKFS_g_s:    Verallgemeinerter und erweiterter Kalman-Filter und Glätter in globaler
              scoring Form.
D-EKFS_g_s:  Wie EKFS_g_s, nur für den diffusen Fall.
Hybrid I:    (IGKFG) (= Fisher-Scoring-Verfahren).
D-Hybrid I:  Wie Hybrid I, nur für den diffusen Fall.

```

- Grafikausgabebildschirm:

Es kann aus 9 verschiedene Grafiken zur Darstellung der Ergebnisse gewählt werden:

- 1.) Die geschätzten $\alpha_{t|\bar{T}}$ mit punktweise Konfidenzbänder
- 2.) $h(Z \mid \alpha_{t|\bar{T}})$ mit punktweise Konfidenzbänder
- 3.) μ_t mit punktweise Konfidenzbänder
- 4.) Die beobachteten y_t
- 5.) Die relativen y_t zusammen mit $h(Z \mid \alpha_{t|\bar{T}})$: Nur für die Multinomialverteilung möglich.
- 6.) y_t zusammen mit μ_t
- 7.) Wie 5.) , nur zusätzlich mit punktweise Konfidenzbänder. Voraussetzung: Multinomialverteilung
- 8.) Wie 6.) , nur zusätzlich mit punktweise Konfidenzbänder
- 9.) Survivor-Funktion ohne Konfidenzbänder
- 0.) Beenden

Belegung des Vektors *index* in PME.ST

Der Vektor *index* wird, wenn z.B. DATEN der Datensatzname ist, mit den Komponenten des Files DATENi.asc belegt. Die nachfolgende Tabelle enthält den Belegungsplan von *index* :

Komponente des Vektors <i>index</i>	Bedeutung
1	T = Anzahl der Beobachtungszeitpunkte
2	\bar{U} = Anzahl der Beobachtungseinheiten bzw. der Gruppen bei gruppierten Daten
3	$\bar{J} = \dim(y_{it})$
4	$n = \dim(\alpha_{t \bar{T}})$
5	<div>Nummer der Response-Funktion h mit</div> <div>0 : identischer link:</div> <div>$h = id$</div> <div>1 : multinomiales logit-Modell:</div> <div>$h_j(\eta) = \frac{\exp(\eta_j)}{1 + \sum_{s=1}^{\bar{J}} \exp(\eta_s)}$</div> <div>2 : $h = \frac{1}{2} + \frac{1}{\pi} \arctan(\eta)$</div> <div>3 : loglineares Poisson-Modell:</div> <div>$h = \exp(\eta)$</div> <div>4 : kumulatives logistic-Modell = proportional odds-Modell (für feste Effekte):</div> <div>$h_j(\eta) = \frac{\exp(\eta_j)}{1 + \exp(\eta_j)} - \frac{\exp(\eta_{j-1})}{1 + \exp(\eta_{j-1})}$</div> <div>5 : grouped Cox-Modell = proportional hazards-Modell (für feste Effekte):</div> <div>$h_j(\eta) = \exp(-\exp(\eta_{j-1})) - \exp(-\exp(\eta_j))$</div> <div>6 : Exponential-Modell:</div> <div>$h_j(\eta) = \exp(-\eta_{j-1}) - \exp(-\eta_j)$</div> <div>7 : grouped proportional hazards-Modell (für feste Effekte):</div> <div>$h_j(\eta) = \frac{\exp(\eta_j)}{\sum_{s=1}^m \exp(\eta_s)} \left\{ 1 - \exp\left(-\sum_{s=1}^{\bar{J}} \exp(\eta_s)\right) \right\}$</div> <div>8 : Probit-Modell:</div> <div>$h = \Phi(\eta)$</div> <div>9 : Numerisch stabiles kumulatives logistic-Modell (siehe Fahrmeir, Tutz (1994), Abschnitt 3.3.3)</div>

6	Nummer der Exponentialfamilie: 1 : Normalverteilung 2 : Poissonverteilung 3: Binomialverteilung 4: scaled Binomialverteilung 5 : Multinomialverteilung 6 : scaled Multinomialverteilung
7	Grafikausgabeschalter: 0 : keine Grafikausgabe 1 : Grafikausgabe
8	Speichermodalitäten zur globalen Variable Vp
9	Speichermodalitäten zur globalen Variable Z
10	Speichermodalitäten zur globalen Variable F
11	Speichermodalitäten zur globalen Variable Q0
12	Speichermodalitäten zur globalen Variable Q
13	Zensierungsmodalitäten: 0 : Zensierung am Anfang des Intervalls 1 : Zensierung am Ende des Intervalls
14	Schalter für Hyperparameterschätzung: 0 : keine Hyperparameterschätzung 1 : Hyperparameterschätzung mit EM- bzw. EM-type Algorithmus

Die anschließenden Übersichten liefern den Schlüssel, mit welcher Zahl die zur jeweiligen globalen Variable gehörige Komponente des Vektors *index* zu belegen ist:

VP (Verteilungsparameter):

VP ist				Belegung
cross-varying	time-varying	diagonal	$\sigma_t^2 \cdot I$	von index (8)
1	1	1	0	1
1	1	0	1	2
1	1	0	0	3
1	0	1	0	4
1	0	0	1	5
1	0	0	0	6
0	1	1	0	7
0	1	0	1	8
0	1	0	0	9
0	0	1	0	10
0	0	0	1	11
0	0	0	0	12

Z (Matrix im Beobachtungsmodell):

Z ist			Belegung
cross-varying	time-varying	diagonal	von index (9)
1	1	1	1
1	1	0	2
1	0	1	3
1	0	0	4
0	1	1	5
0	1	0	6
0	0	1	7
0	0	0	8

F (Übergangsmatrix im Übergangsmodell)

F ist		Belegung
time-varying	diagonal	von index (10)
1	1	1
1	0	2
0	1	3
0	0	4

Q (Kovarianzmatrix im Übergangsmodell)

Q ist		Belegung
time-varying	diagonal	von index (12)
1	1	1
1	0	2
0	1	3
0	0	4

Q0 (Kovarianzmatrix von β_0 im Übergangsmodell)

Belegung von index (11): 0 - vollbesetzt 1 - diagonal

B Beweise

Kapitel 2, Formel (2.4):

Beh.: $E(T) = \int_0^b S(t) dt.$

Bew.: Da $T \geq 0$ ist, gilt $F(t) = 0$ für $t \leq 0$.

Ferner sei $b \in \mathbb{R}$ und $b > 0$ so, daß $F(b) = 1$.

Damit:

$$\begin{aligned}
 E(T) & \stackrel{\text{Stieltjes-Integral}}{=} \int_0^b t dF(t) = \int_0^b t \cdot f(t) dt \stackrel{\text{partielle Integration}}{=} [t \cdot F(t)]_0^b - \int_0^b F(t) dt = \\
 & = b \cdot F(b) - \int_0^b F(t) dt = b - \int_0^b F(t) dt = \int_0^b 1 dt - \int_0^b F(t) dt = \int_0^b \{1 - F(t)\} dt = \\
 & = \int_0^b S(t) dt. \qquad \qquad \qquad \text{q.e.d}
 \end{aligned}$$

Kapitel 2, Formel (2.7):

Beh.: $S(t) = \prod_{i=1}^{t-1} (1 - \lambda(i)).$

Bew.: Die diskrete Zufallsvariable T nimmt Werte $t \in \{1, \dots, \bar{T}\}$ an.

Damit gilt $\text{pr}(T \geq 1) = 1$. Das benutzen wir in

$$\begin{aligned}
 S(t) = \text{pr}(T \geq t) &= \frac{\text{pr}(T \geq t)}{1} = \frac{\text{pr}(T \geq t)}{\text{pr}(T \geq 1)} = \frac{\text{pr}(T > t-1)}{\text{pr}(T > 0)} = \\
 &= \prod_{i=1}^{t-1} \left(\frac{\text{pr}(T > i)}{\text{pr}(T > i-1)} \right) = \prod_{i=1}^{t-1} \left(\frac{\text{pr}(T \geq i) - \text{pr}(T = i)}{\text{pr}(T \geq i)} \right) =
 \end{aligned}$$

$$= \prod_{i=1}^{t-1} \left(1 - \frac{\text{pr}(T=i)}{\text{pr}(T \geq i)} \right) = \prod_{i=1}^{t-1} (1 - \text{pr}(T=i | T \geq i)) = \prod_{i=1}^{t-1} (1 - \lambda(i)) . \quad \text{q.e.d.}$$

Kapitel 2, Formel (2.11):

Beh.: $f_j(t) = \lambda_j(t) \cdot S(t)$.

$$\text{Bew.: } \lambda_j(t) \stackrel{(2.8)}{=} \text{pr}(T=t, J=j | T \geq t) \stackrel{\text{bedingte Wahrscheinlichkeit}}{=} \frac{\text{pr}(T=t, J=j)}{\text{pr}(T \geq t)} = \frac{f_j(t)}{S(t)} .$$

Daraus ergibt sich für $S(t) \neq 0$ die Behauptung. q.e.d.

Kapitel 2, Formel (2.12):

Beh.: $\text{pr}(J=j) = \sum_{t=1}^{\bar{T}} \lambda_j(t) S(t)$.

Bew.: Da die diskrete Zufallsvariable T Werte $t \in \{1, \dots, \bar{T}\}$ annimmt, gilt:

$$\text{pr}(J=j) = \sum_{t=1}^{\bar{T}} \text{pr}(J=j, T=t) \stackrel{(2.11)}{=} \sum_{t=1}^{\bar{T}} f_j(t) \stackrel{(2.11)}{=} \sum_{t=1}^{\bar{T}} \lambda_j(t) S(t) . \quad \text{q.e.d.}$$

Kapitel 3, Formel (3.14):

Beh.: h in (3.14) ist die Inverse zur natürlichen Linkfunktion der Multinomialverteilung.

Bew.: Zur Vereinfachung der Notation werden im folgenden die Indizes t für die Zeit und U für die Untersuchungseinheit nicht angegeben.

Mit den Parametern der Multinomialverteilung $\pi^k := (\pi_1^k, \dots, \pi_J^k)'$ und $g(\pi^k) := (g_1(\pi^k), \dots, g_J(\pi^k))'$ ist nach Fahrmeir und Tutz (1994, Formel (3.1.7)) die natürliche Linkfunktion der Multinomialverteilung $g(\pi^k)$ gegeben durch:

$$g_j(\pi^k) = \ln \left(\frac{\pi_j^k}{1 - \sum_{i=1}^J \pi_i^k} \right) .$$

Da g eine Linkfunktion ist, gilt

$$\ln \left(\frac{\pi_j^k}{1 - \sum_{i=1}^J \pi_i^k} \right) = \eta_j^k, \quad (\text{B.1})$$

wobei η_j^k der lineare Prädiktor wie in Abschnitt 3.1. Nach (B.1) gilt

$$\begin{aligned} \sum_{j=1}^J \exp(\eta_j^k) &= \sum_j \frac{\pi_j^k}{1 - \sum_i \pi_i^k} = \frac{\sum_j \pi_j^k}{1 - \sum_i \pi_i^k} = \frac{1}{\frac{1}{\sum_j \pi_j^k} - 1} \Rightarrow \\ \Rightarrow \frac{1}{\sum_j \pi_j^k} - 1 &= \frac{1}{\sum_j \exp(\eta_j^k)} \Rightarrow \\ \Rightarrow \frac{1}{\sum_j \pi_j^k} &= \frac{1}{\sum_j \exp(\eta_j^k)} + 1 = \frac{1 + \sum_j \exp(\eta_j^k)}{\sum_j \exp(\eta_j^k)} \Rightarrow \\ \Rightarrow \sum_j \pi_j^k &= \frac{\sum_j \exp(\eta_j^k)}{1 + \sum_j \exp(\eta_j^k)}. \end{aligned} \quad (\text{B.2})$$

Ferner gilt nach (B.1):

$$\begin{aligned} \exp(\eta_j^k) &= \frac{\pi_j^k}{1 - \sum_i \pi_i^k} \Rightarrow \\ \Rightarrow \pi_j^k &= \exp(\eta_j^k) \cdot \left(1 - \sum_i \pi_i^k \right) \stackrel{(\text{B.2})}{=} \exp(\eta_j^k) \cdot \left(1 - \frac{\sum_i \exp(\eta_i^k)}{1 + \sum_i \exp(\eta_i^k)} \right) = \\ &= \exp(\eta_j^k) \cdot \left(\frac{1}{1 + \sum_i \exp(\eta_i^k)} \right) = \frac{\exp(\eta_j^k)}{1 + \sum_i \exp(\eta_i^k)}. \end{aligned}$$

Mit $h_j(\eta^k) = \pi_j^k$ erhält man die natürliche Responsefunktion wie in (3.14).

q.e.d.

Kapitel 3, Formel (3.19):

Beh.: $\text{pr}(T \in [b_{t-1}, b_t)) = S(b_t)[1 - \exp(w_t \cdot \lambda_t)]$, wobei $S(t)$ die Survivor-Funktion aus (2.1) ist.

$$\begin{aligned}
 \text{Bew.: } \text{pr}(T \in [b_{t-1}, b_t)) &= S(b_{t-1}) - S(b_t) \stackrel{(2.3)}{=} \exp\left(-\int_0^{b_{t-1}} \lambda(t) dt\right) - \exp\left(-\int_0^{b_t} \lambda(t) dt\right) = \\
 &= \exp\left(-\int_0^{b_{t-1}} \lambda(t) dt\right) - \exp\left\{-\left(\int_0^{b_{t-1}} \lambda(t) dt + \int_{b_{t-1}}^{b_t} \lambda(t) dt\right)\right\} = \\
 &= \exp\left(-\int_0^{b_{t-1}} \lambda(t) dt\right) \cdot \left\{-\exp\left(-\int_{b_{t-1}}^{b_t} \lambda(t) dt\right) + 1\right\} \stackrel{\text{mit (2.3) und (3.18)}}{=} \\
 &= S(b_{t-1}) \cdot \{1 - \exp(w_t \cdot \lambda_t)\}. \quad \text{q.e.d.}
 \end{aligned}$$

Kapitel 5, Formel (5.28):

Beh.: $V_{t|t} = (I_n + V_{t|t-1} S_t(\alpha_t))^{-1} V_{t|t-1}$.

Bew.: Im Korrekturschritt von (D-EKFG), (D-LGKFG), (EKFG) und (LGKFG) gilt:

$$V_{t|t} = \{V_{t|t-1}^{-1} + S_t(\alpha_t)\}^{-1} \quad (\text{B.3})$$

Es gilt, wobei wir stets Invertierbarkeit voraussetzen:

$$\begin{aligned}
 (\text{B.3}) &\Leftrightarrow \{(V_{t|t-1}^{-1} + S_t(\alpha_t))V_{t|t} = I_n\} \Leftrightarrow \{V_{t|t-1}^{-1}V_{t|t} + S_t(\alpha_t)V_{t|t} = I_n\} \Leftrightarrow \\
 &\Leftrightarrow \{V_{t|t} + V_{t|t-1}S_t(\alpha_t)V_{t|t} = V_{t|t-1}\} \Leftrightarrow \{(I_n + V_{t|t-1}S_t(\alpha_t))V_{t|t} = V_{t|t-1}\} \Leftrightarrow \\
 &\Leftrightarrow \{V_{t|t} = (I_n + V_{t|t-1}S_t(\alpha_t))^{-1}V_{t|t-1}\}. \quad \text{q.e.d.}
 \end{aligned}$$

Literaturverzeichnis

- Anderson, B.D.O. und Moore, J.B. (1979): *Optimal Filtering*, Prentice-Hall, Englewood Cliffs, New Jersey.
- Arjas, E. (1981): The failure and hazard processes in multivariate reliability systems, *Math. Oper. Res.* **6**, 551-562.
- Arjas, E. (1989): Survival Models and Martingale Dynamics, *Scand. J. Statist.* **16**, 177-225.
- Aydemir, S. (1992): Dynamische Modelle zur Ereignisanalyse, Diplomarbeit, Institut für Statistik, Ludwig-Maximilians-Universität München.
- Blossfeld, H.-P., Hamerle, A. und Mayer, K.U. (1986): *Ereignisanalyse: statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften*, Campus Verlag, Frankfurt am Main, New York.
- Carlin, B.P., Polson, N.G. und Stoffer, D.S. (1992): A Monte Carlo approach to nonnormal and nonlinear state-space modeling, *JASA* **87**, 493-500.
- Çınlar, E. (1972): Markov additive processes. II, *Z. Wahrsch. Verw. Gebiete* **24**, 94-121.
- Courgeau, D. und Lelièvre, É. (1992): *Event History Analysis in Demography*, Clarendon Press, Oxford.
- Cox, D.R. und Oakes, D. (1984): *Analysis of Survival Data*, Chapman and Hall, London, New York.
- De Jong, P. (1989): Smoothing and interpolation with the state space model, *JASA* **84**, 1085-1089.

- De Jong, P. (1991): The diffuse Kalman Filter, *The Annals of Statistics* **19**, 1073-1083.
- De Jong, P. und Mackinnon, M.J. (1988): Covariances for smoothed estimates in state space models, *Biometrika* **75**, 601-602.
- Deutsches Institut für Wirtschaftsforschung (Hrsg.) (1993): Das Sozio-oekonomische Panel (SOEP), Benutzerhandbuch Band II Version 7 - November/93, Berlin.
- Durbin, J. und Koopman, S.J. (1992): Kalman filtering and smooting for non-Gaussian time series, Discussion paper, London School of Economics and Political Science.
- Efron, B. (1988): Logistic regression, survival analysis, and the Kaplan-Meier Curve, *JASA* **83**, 414-425.
- Fahrmeir, L. (1992): Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models, *JASA* **87**, 501-509.
- Fahrmeir, L. (1994): Dynamic modelling and penalized likelihood estimation for discrete time survival data, *Biometrika* **81**, 317-330.
- Fahrmeir, L. und Hamerle, A., Hrsg. (1984): *Multivariate statistische Verfahren*, Walter de Gruyter, Berlin, New York.
- Fahrmeir, L. und Kaufmann, H. (1991): On Kalman Filtering, Posterior Mode Estimation and Fisher Scoring in Dynamic Exponential Family Regression, *Metrika* **38**, 37-60.
- Fahrmeir, L. und Tutz, G. (1994): *Multivariate Statistical Modelling Based on Generalized Linear Models*, Springer, New York.
- Fahrmeir, L. und Wagenpfeil, S. (1995a): Penalized Likelihood Estimation and Iterative Kalman Smoothing for Non-Gaussian Dynamic Regression Models, Discussion Paper Nr. 5, SFB 386, Ludwig-Maximilians-Universität, München.

- Fahrmeir, L. und Wagenpfeil, S. (1995b): Smoothing Hazard Functions and Time-Varying Effects in Discrete Duration and Competing Risks Models, Discussion Paper Nr. 7, SFB 386, Ludwig-Maximilians-Universität, München.
- Fahrmeir, L., Kaufmann, H.L. und Ost, F. (1981): *Stochastische Prozesse. Eine Einführung in Theorie und Anwendungen*, Carl Hanser Verlag, München, Wien.
- Fleming, Th.R. und Harrington, D.P. (1991): *Counting Processes and Survival Analysis*, John Wiley & Sons, Inc., New York, Chichester, Brisbane, Toronto, Singapore.
- Friedman, M. (1982): Piecewise exponential models for survival data with covariates, *The Annals of Statistics* **10**, 101-113.
- Gamerman, D. (1991): Dynamic Bayesian models for survival data, *Appl. Statist.* **40**, 63-79.
- Gamerman, D. (1992): A dynamic approach to the statistical analysis of point processes, *Biometrika* **79**, 39-50.
- Gamerman, D. und West, M. (1987): An application of dynamic survival models in unemployment studies, *The Statistician* **36**, 269-274.
- Gill, P.E., Murray, W. und Wright, M.H. (1981): *Practical Optimization*, Academic Press, London, San Diego, New York.
- Härdle, W. (1993): *Applied Nonparametric Regression*, Reprint, Cambridge University Press, Cambridge.
- Hamerle, A. (1985a): Zählprozeß-Modelle zur statistischen Analyse von Ereignisdaten mit Kovariablen bei konkurrierenden Risiken und mehreren Episoden, Diskussionsbeiträge Nr. 90/s, Fakultät für Wirtschaftswissenschaften und Statistik der Universität Konstanz, Konstanz.
- Hamerle, A. (1985b): Regressionsmodelle für gruppierte Verweildauern und Lebenszeiten, *ZOR-Zeitschrift Operations Research* **29 B**, 243-260.

- Hamerle, A. (1986): Regression analysis for discrete event history or failure time data, *Statistical Papers* **27**, 207-225.
- Hamerle, A. (1989): Multiple-spell Regression Models for Duration Data, *Appl. Statist.* **38**, 127-138.
- Hamerle, A. und Tutz, G. (1989): *Diskrete Modelle zur Analyse von Verweildauer und Lebenszeit*, Campus Verlag, Frankfurt am Main, New York.
- Harvey, A.C. (1992): *Forecasting, structural time series models and the Kalman filter*, Reprint, Cambridge University Press, Cambridge.
- Kalbfleisch, J.D. und Prentice, R.L. (1980): *The Statistical Analysis of Failure Time Data*, John Wiley & Sons, New York, Chichester.
- Kass, R.E. und Raftery, A.E. (1994): Bayes Factors, Technischer Bericht Nr. 254, Department of Statistics, University of Washington, Technischer Bericht Nr. 571, Department of Statistics, Carnegie-Mellon-University.
- Kendall, M., Stuart, A. und Ord, J.K. (1987): *Kendall's Advanced theory of statistics*, 5. Ausgabe, Band 1: Distribution Theory, Charles Griffin & Company Limited, London.
- Knorr-Held, L. (1995): Markov Chain Monte Carlo Simulation in Dynamic Generalized Linear Mixed Models, Discussion Paper Nr. 8, Sonderforschungsbereich 386 der Ludwig-Maximilians-Universität München, München.
- Koopman, S.J. (1993): Disturbance Smoother for state space models, *Biometrika* **80**, 117-126.
- Kredler, Ch. (1986): Behaviour of third order terms in quadratic approximations of LR-statistics in multivariate generalized linear models, *The Annals of Statistics* **14**, 326-335.

- Lancaster, T. (1990): *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge, New York.
- Lawless, J.F. (1982): *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York, Chichester.
- Lemoine, A.J. and Wenocur, M.L. (1985): On failure modeling, *Naval Res. Logist. Quart.* **32**, 497-508.
- Lindsey, J.K. (1993): *Models for Repeated Measurements*, Clarendon Press, Oxford.
- Narendranathan, W. und Stewart, M.B. (1991): Simple Models for Testing for the Proportionality of Cause-Specific Hazards in Competing Risks Models, *Oxford Bulletin of Economics and Statistics* **53**, 331-340.
- Narendranathan, W. und Stewart, M.B. (1993): Modelling the Probability of Leaving Unemployment: Competing Risks Models with Flexible Base-Line Hazards, *Appl. Statist.* **42**, 63-83.
- Projektgruppe Sozio-ökonomisches Panel (1995): Das Sozio-oekonomische Panel (SOEP) im Jahre 1994, *Vierteljahreshefte zur Wirtschaftsforschung* **1**, 5-15.
- Pudney, S. und Thomas, J. (1995): Specification Tests for the Competing Risks Duration Model: An Application to Unemployment Duration and Sectoral Movement, *Oxford Bulletin of Economics and Statistics* **57**, 323-348.
- Singpurwalla, N.D. (1995): Survival in Dynamic Environments, *Statistical Sciences* **10**, 86-103.
- Snyder, D.L. (1975): *Random Point Processes*, John Wiley & Sons, New York.

- Tierney, L. (1994): Markov chains for exploring posterior distributions, *Ann. Statist.* **22**, 1701-1762.
- Ulm, K., Schmoor, C., Sauerbrei, W., Kemmler, G., Aydemir, Ü., Müller, B. und Schumacher, M. (1989): Strategien zur Auswertung einer Therapiestudie mit der Überlebenszeit als Zielkriterium, *Biometrie und Informatik in Medizin und Biologie* **20**, 171-205.
- Wagenpfeil, S. (1995): Hyperparameter estimation in exponential family state space models, Discussion Paper Nr. 6, SFB 386, Ludwig-Maximilians-Universität, München..
- Wenocur, M.L. (1989): A reliability model based on the gamma process and its analytic theory, *Adv. in Appl. Probab.* **21**, 899-918.
- Werner, J. (1992): *Numerische Mathematik 2*, Vieweg Verlag, Braunschweig, Wiesbaden.
- West, M. und Harrison, P.J. (1989): *Bayesian Forecasting and Dynamic Models*, Springer-Verlag, New York.
- West, M., Harrison, P.J. and Migon, H.S. (1985): Dynamic generalized linear models and Bayesian forecasting, *JASA* **80**, 73-83.

Ehrenwörtliche Versicherung

Ich versichere hiermit ehrenwörtlich, daß ich die Arbeit selbstständig und ohne fremde Hilfe verfaßt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie mir gegebene Anregungen sind als solche kenntlich gemacht.

Die Arbeit wurde bisher noch keiner anderen Prüfungsbehörde vorgelegt und noch nicht veröffentlicht.

München, den 6. Oktober 1995

Stefan Wagenpfeil